

VU Research Portal

Developmental timing and cell fate maintenance in *Caenorhabditis elegans*

Traets, Joleen Johanna Helena

2021

document version

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

citation for published version (APA)

Traets, J. J. H. (2021). *Developmental timing and cell fate maintenance in Caenorhabditis elegans*. [PhD-Thesis - Research and graduation internal, Vrije Universiteit Amsterdam].

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

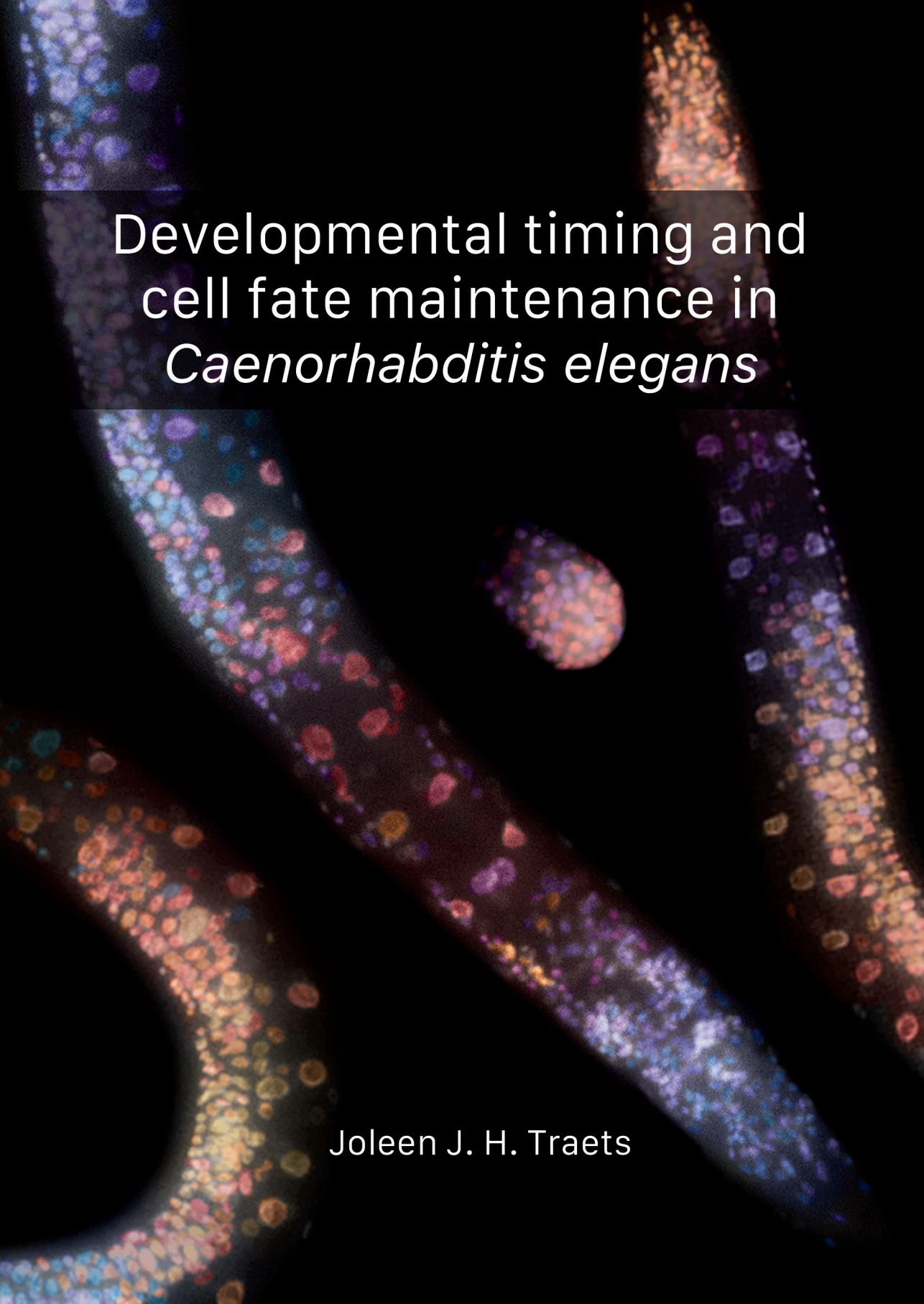
- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

vuresearchportal.ub@vu.nl

A fluorescence microscopy image of a C. elegans germline. The image shows several long, curved structures (germline arms) filled with numerous small, brightly colored fluorescent spots. These spots are primarily blue, purple, and red, with some yellow/orange spots visible in the lower left. The background is dark, making the fluorescent structures stand out. The text is overlaid on the upper part of the image.

Developmental timing and cell fate maintenance in *Caenorhabditis elegans*

Joleen J. H. Traets

**Developmental timing and cell fate
maintenance in *Caenorhabditis elegans***

Joleen J.H. Traets

Members of the thesis committee:

prof. dr. ir. E.J.G. Peterman	Vrije Universiteit Amsterdam
prof. dr. P.R. ten Wolde	AMOLF / Vrije Universiteit Amsterdam
prof.dr. H.C. Korswagen	Hubrecht Institute
dr. L. Laan	TU Delft
dr. S. Semrau	Leiden University

PhD thesis, Vrije Universiteit Amsterdam, 10th of February, 2021

Title: Developmental timing and cell fate maintenance *Caenorhabditis elegans*

Author: Joleen J.H. Traets

Cover: “Een ode aan alle gesneuvelde wormen”, Joleen J.H. Traets

ISBN: 978-94-92323-47-7

Copyright © 2021 by Joleen J.H. Traets.

The work described in this dissertation is part of the research program of the Foundation for Fundamental Research on Matter (FOM), which is financially supported by the Netherlands Organization for Scientific Research (NWO).



The work described in this thesis was performed at AMOLF, Science Park 104, 1098 XG Amsterdam, The Netherlands.

An electronic version of this dissertation is available at <http://www.amolf.nl> and at <http://www.vu.nl>

VRIJE UNIVERSITEIT

Developmental timing and cell fate maintenance in *Caenorhabditis elegans*

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor of Philosophy aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. V. Subramaniam,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Bètawetenschappen
op woensdag 10 februari 2021 om 15.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door

Joleen Johanna Helena Traets

geboren te Roosendaal en Nispen

promotor: prof.dr. T.S. Shimizu

copromotor: dr. J.S. van Zon

Table of contents

General introduction	1
1.1 Timing in development	4
1.2 Maintenance of terminal cell fate	5
1.3 Genetic networks underlying cell fate switches	6
1.4 Stochasticity in molecular networks of developmental processes	8
1.5 <i>C. elegans</i> as a model system	9
1.6 Terminal cell fate maintenance of ASE neurons	11
1.7 Oscillating genes in the pharynx	12
1.9 References	16
 WormFISH: a software package for single molecule FISH analysis of <i>C. elegans</i> larval development	 21
2.1 Abstract	21
2.2 Introduction	23
2.3 Results	25
2.3.1 Overview of smFISH protocol	25
2.3.2 Workflow for identifying and counting smFISH spots	27
2.3.3 Annotation of developmental age in fixed animals	29
2.3.4 User interface and workflow of the WormFISH package	32
2.3.5 Quality of the fluorescent spot identification	34
2.4 Discussion	35
2.5 Material and methods	37
2.6 Appendix	38
2.6.1 Author contributions to the chapter	38
2.6.2 Supplementary figures	38
2.7 References	39
 Mechanism of life-long maintenance of neuron identity despite molecular fluctuations	 43
3.1 Abstract	43
3.2 Introduction	45
3.3 Results	47
3.3.1 Loss of ASE neuron fate upon transient CHE-1 depletion	47
3.3.2 Copy number and lifetime of <i>che-1</i> mRNA and protein	49
3.3.3 Stochastic simulations identify stable cell fate maintenance parameters	52
3.3.4 Stability against stochastic fluctuations by preferential binding of CHE-1 to its own promoter	55
3.3.5 <i>In vivo</i> CHE-1 depletion decreases target gene but not <i>che-1</i> expression	56
3.3.6 Sequences flanking the <i>che-1</i> ASE motif are required for <i>che-1</i> expression during CHE-1 depletion	57

3.3.7 Involvement of an <i>Otx</i> -related homeodomain binding site in maintaining ASE subtype	59
3.4 Discussion	61
3.5 Material and methods	64
3.6 Appendix	75
3.6.1 Acknowledgements	75
3.6.2 Author contributions	75
3.6.3 Supplementary figures	76
3.7 References	81
Validation of pharynx and neuron related clusters found in RNA-sequencing data with novel density-based clustering approach	85
4.1 Abstract	85
4.2 Introduction	87
4.3 Results	89
4.3.1 Choice of genes representative for two clusters found in RNA-sequencing data with novel density-based clustering approach	89
4.3.2 Highly specific expression pattern and biological functions of cluster genes	91
4.3.3 Spatial expression patterns and correlations between cluster genes	93
4.3.3.1 Analysis of the neuron cluster genes	93
4.3.3.2 Analysis of the pharynx cluster genes	96
4.3.4 Transgenerational memory effect on starvation cannot explain the correlations within or between the clusters	99
4.3.5 Direct starvation response induces correlation between clusters due to age related upregulation of gene expression	101
4.4 Discussion	103
4.5 Material and methods	106
4.6 Appendix	108
4.6.1 Author contribution to the chapter	108
4.6.2 Supplementary figures	109
4.7 References	114
Oscillating gene expression of the molting cycle oscillator is coordinated with pharynx growth in larvae	119
5.1 Abstract	119
5.2 Introduction	121
5.3 Results	123
5.3.2 Oscillating genes in the pharynx region	123
5.3.2 Oscillating expression levels of genes expressed in different pharyngeal cell types	125
5.3.2 Growth of the pharynx is coordinated with oscillatory gene expression	128
5.3.3 Arrested L1 larvae halt pharynx growth and gene expression oscillations	131
5.3.4 Arrested gene expression oscillations in late-larval developmental	

arrest	132
5.3.5 Variation in pharynx/body size ratio in late-larval arrests	135
5.3.6 Variability in arrest of pharynx growth in starving animals	135
5.3.7 Variability in gene expression levels of oscillating genes in animals upon starvation	137
5.4 Discussion	139
5.5 Material and methods	141
5.6 Appendix	144
5.6.1 Author contribution to the chapter	144
5.6.2 Supplementary figures	145
5.7 References	148
General discussion & outlook	153
6.1 Long-term maintenance of ASE fate with the CHE-1 switch	156
6.2 Growth of the pharynx coordinated with molting cycle oscillator	159
6.3 References	163
Appendix	167
7.1 Summary	167
7.2 Samenvatting	171
7.3 Acknowledgements	177

Chapter 1

General introduction

The development of a multi-cellular organism begins with a single, fertilized egg cell and ends with a fully-developed organism composed of multiple kinds of specialized cell types. During the process of development, cells undergo several important developmental events, including specification, differentiation, proliferation, movement of cells and growth (Fig. 1). The correct execution of those events comes with many challenges. One key challenge is the correct timing of the occurrence and duration of developmental events, with errors in timing potentially leading to malformation and dysfunction of tissue that could be incompatible with survival of the organism. When cells are differentiated and have reached their specialized terminal cell fate, a new challenge appears, which is maintaining the obtained terminal cell fate to prevent loss of function of the specialized cell types. In this thesis, I will focus on these two challenges that cells face during their lifetime: the timing of developmental events, and the maintenance of the terminal cell fate. In the following section of the introduction, I will first describe the general knowledge that has been gathered over the past decades and elaborate more on the importance of both aspects.

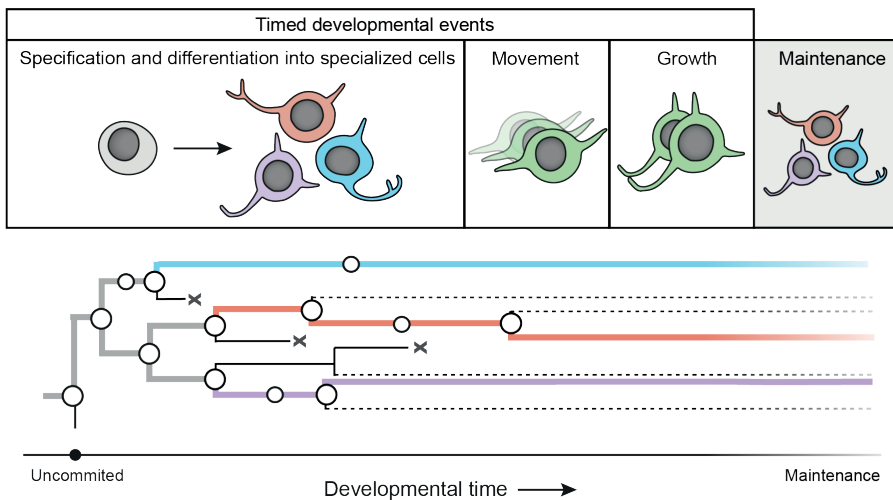


Figure 1: Developmental timing and maintenance. During development of an uncommitted cell to a specialized cell type, cells undergo several developmental events. For example, uncommitted cells specify and differentiate into different specialized cell types, undergo movement and growth as well as proliferation and cell death. Finally, when the cells have reached their specialized cell fate, the identity has to be maintained. These events occur in a specific order and at a specific time on the developmental time line of the organism. Both developmental timing and maintenance must be executed correctly to ensure correct organ or tissue structure. Coloured lines: represented the three specialized cell types in developmental trajectory. Large circles: developmental events related to growth and movement. Small circles: developmental events related to specification and differentiation. Dotted lines: other cells that have reached terminal cell identity. X: cell death.

1.1 Timing in development

The complex developmental process of a fertilized egg cell to a multicellular organism consists out of series of sequential developmental events. An example of a sequence of developmental events is one in which cells are first specified and then determined to their committed cell fate, after which cells differentiate into their terminal cell fate. Other important developmental events, include the movement of cells to the appropriate location, cell divisions and cell death. Also, growth of cell size can be seen as a developmental event, in which cells increase their volume at a certain time in development. The combination of sequential developmental events leads up to the formation of, e.g. tissues and complex organs. It is of particular importance that the order and duration of the sequentially timed developmental events is executed correctly. Some variation in timing can produce the subtle differences between individuals, but major deviation in timing might lead to, for example malformation of tissues that are incompatible with survival.

In general, the timing of developmental events occurs without the influence of external time, meaning that the schedule of developmental related events must be generated internally. In some cases, developmental timing appears to be generated in a fully cell-autonomous fashion (Wolpert 2015). For example, during neuronal development in *Drosophila* the difference in the time of cell birth determines the neuronal subtypes generate in each division (Toma, Wang, and Hanashima 2016). In this case, several transcription factors are sequentially activated and repressed in turn, inducing the execution of the developmental related events in the correct order. However, correct timing of development also likely relies on external timing cues that impact multiple cells and even the entire organism. Moreover, while the above example explains how the correct sequence of cell fate decisions is established, it cannot explain the overall pace at which development progresses. To regulate the pace of development and ensure that parallel developmental processes remain synchronized, likely requires measuring time on the level of the organism (Ebisuya and Briscoe 2018).

One way to keep track of time is by biological timers. Those timers can have gradual increasing or decreasing expression levels in time, in which a threshold determines the activation of the developmental related events (Ebisuya and Briscoe 2018). Another timing mechanism is the well-described segmentation clock, in which oscillatory gene expression is utilized to drive the periodic developmental events related to the generation of somites in embryos (Kageyama et al. 2012; Oates, Morelli, and Ares 2012). Other periodic developmental events thought to be driven by a similar mechanism of gene expression oscillations, is the molting cycle in *Caenorhabditis elegans*, involving the recurring event of renewal and shedding

of the skin (Hendriks et al. 2014; Kim, Grun, and van Oudenaarden 2013; Turek and Bringmann 2014). However, even though gene expression oscillators could be a powerful mechanism to control timing of development, it is currently not understood how these oscillations are generated and controlled, and how these oscillations regulate the timing of downstream developmental processes.

1.2 Maintenance of terminal cell fate

During terminal differentiation, the precursor cell loses its ability to proliferate and becomes irreversibly a specialized cell type. Those terminally differentiated cells express a battery of cell-specific genes that are characteristic for the specific cell and required for their specialized function. Importantly, the specialized function of the terminally differentiated cells needs to be maintained during the life-span of the cells. Defects in terminal cell fate maintenance can lead to tissue malfunction or induce tumour formation (Wolpert 2015). An important paradigm for cell fate maintenance is the concept of “terminal selector” genes (Hobert 2008, 2016), specialized transcription factors that are solely responsible for the expression of a battery of cell-specific genes, by binding on specific cis-regulatory target sites. This maintenance function of the terminal selector genes is mediated by inducing their own expression via positive autoregulation (Hobert 2008). Another important mechanism is the altering of expression of cell-specific genes by remodelling the chromatin structure, e.g. by making the chromatin structure (in)accessible for gene transcription (Wolpert 2015). This mechanism is known to be particularly important for prevention of induction of unwanted terminal cell fate in an already terminally differentiated cell.

Some cells are renewed every few days, such as epithelial cells in the gastrointestinal tract with a turnover between 2-10 days (Darwich et al. 2014; Spalding et al. 2005). In contrast, most neuronal cells are not renewed and have to maintain their terminal cell identity for the entire lifetime of an organism (Ming and Song 2005; Spalding et al. 2005). The stable maintenance of terminal cell identity is particularly important in those long-lived cells to prevent premature loss of function. The terminal selector mechanism is found to underlie terminal differentiation of neuronal cells in several organisms, and has so far been identified in various organisms, including *C. elegans*, *Drosophila melanogaster* and *Mus musculus* (Hsiao et al. 2013; Konstantinides et al. 2018; Hirota and Mombaerts 2004; Gabilondo et al. 2016). This broad occurrence of terminal selector genes suggests that this principle for terminal differentiation is evolutionary conserved. However, the role of the terminal selectors and their potential role in the long-term maintenance of the cell fate, specifically in long-lived cells, has so far been poorly studied.

In this thesis, I will focus on these complementary challenges of correct developmental timing and long-term cell fate maintenance that cells face during their lifetime. In particular, I will study how cell growth is timed by gene expression oscillators during development and how long-term maintenance of terminal cell fate is controlled by terminal selector genes. Perhaps surprisingly, the apparently complex process of cell differentiation can be explained by the action of relatively simple genetic networks. In the following section, I will describe in more detail the key concepts behind the networks that control cell differentiation and maintenance, and the impact on their behaviour of intrinsic variability in the level of their components. All the work in this thesis is performed in the nematode worm *C. elegans*. In the final sections, I will introduce this important model organism for developmental biology and explain why it is an ideal model system to study developmental timing and cell fate maintenance. nents. All the work in this thesis is performed in the nematode worm *C. elegans*. In the final sections, I will introduce this important model organism for developmental biology and explain why it is an ideal model system to study developmental timing and cell fate maintenance.

1.3 Genetic networks underlying cell fate switches

In general, the complex behaviour of biological systems can be described with relatively simple genetic regulatory networks consisting of feedback and feed-forward loops (Alon 2007; Tyson and Novak 2010). This also holds for the terminal differentiation and maintenance of terminal cell fate by terminal selectors. Specifically, networks underlying cell fate maintenance appear to function as genetic switches.

In biology, genetic switches play an important role in a wide variety of processes, including the cell cycle, cell fate decisions and apoptosis (Tyson, Chen, and Novak 2003; Yao et al. 2008; Xiong and Ferrell 2003; Ferrell and Machleder 1998; Eissing et al. 2004). The genetic switch has two stable states, called bistability, meaning that theoretically it is reversible and has the potential to make state transitions between both stable states. A classic, well-studied example of a genetic switch is the lambda phage switch (Ptashne and Mark 1992). This switch has two stable states. The first state is called the lysogenic state where the lambda bacteriophage is integrated in the host genome but is otherwise quiescent. The second state is called the lytic state where the lambda bacteriophage replicates in a fast pace causing lysis of the host and allowing the bacteriophages to escape the host. Upon infection, phage typically enter the lysogenic state and can remain in this state for prolonged periods of time. However, a small dose of UV light causes a transient upregulation of the expression of a gene responsible for inducing lysis, with the gene positively autoregulating its

own expression in a positive feedback loop. This results in the switch transitioning to the lytic state, eventually leading to virus production and lysis of the host cell.

In case of cell fate decisions, positive and negative feedback loops are employed to establish and stabilize new cellular states, which, generally, is biologically an irreversible process (Jukam and Desplan 2010). In those systems, an all-or-none decision is made, causing the cell to differentiate into a specialized cell fate. A basic feature of these bistable systems is that once an input signal reaches a defined threshold, the system will switch to the other steady state, in this case the specialized cell fate. In general, this decision is induced by other activators, working only transiently, that trigger the state transition. The system can potentially remain in this steady state long after the signal is removed. There are theoretically many genetic regulatory network architectures that can lead to the described bistability (Ferrell 2002; Ferrell and Xiong 2001). A single positive feedback loop on a terminal selector gene inducing its own expression is already sufficient to generate the described switch-like behavior. Other simple architectures giving rise to bistable behavior is a system in which two genes mutually inhibit each other, called bistable toggle switches (Gardner, Cantor, and Collins 2000).

A well-studied example of a bistable genetic cell fate switch, is the irreversible transition in *Xenopus* oocyte maturation, in which the oocytes can reside in the arrest state for extended periods of time, but only transiently in the intermediate states before transitioning into the matured state (Ferrell and Machleder 1998; Xiong and Ferrell 2003). Many biological systems have not only a single but multiple positive feedback loops (Brandman et al. 2005), as is the case in *Xenopus* oocyte maturation. Moreover, the complex genetic regulatory network underlying the maturation in the oocyte is generated by responding to maturation-inducing stimuli, which activates rapid (de)phosphorylation-mediated positive feedback loops, involving the mitogen-activated protein kinase (MAPK) cascade (Ferrell et al. 2009; Abrieu, Doree, and Fisher 2001).

The architecture of the genetic regulatory networks underlying the differentiation and maintenance of terminal cell fate appears capable of bistable switch-like dynamics. In this case, one steady state of the genetic switch represents the cell when it has not yet adapted the terminal cell fate, and the second steady state represents the terminal cell fate identity of the cell. The genetic regulatory network of terminal selector genes often consists out of positive feedback loops, such as positive autoregulation of the terminal selector genes. An example of such a genetic regulatory network, are the homeodomain proteins UNC-86 and MEC-3 in the touch sensory neuron in *C. elegans*, which are terminal selector genes and act cooperatively as a dimer inducing expression of cell-specific genes (Duggan, Ma, and Chalfie 1998; Xue, Tu,

and Chalfie 1993). Both UNC-86 and MEC-3 are thought to autoregulate their expression as dimers. A simpler architecture of a genetic regulatory network involving terminal selector genes, which is required for the terminal cell fate induction and also maintenance, is the zinc-finger transcription factor *che-1* in the ASE neurons (Etchberger et al. 2007; Leyva-Diaz and Hobert 2019). Here, CHE-1 positively autoregulates its own expression, while it is turned on by Tailless/TLX transcription factor NHR-67 that transiently induces the expression of *che-1* to establish the ASE cell fate.

The loss of terminal cell fate during the life-span of a cell is unwanted. Hence, the terminal cell fate has to be stably maintained by the terminal cell fate switch over extended periods of time. However, bistable systems are inherently reversible, suggesting that there is a potential of losing terminal cell fate if the system is not capable to remain in the steady state for long enough. In the next section, I will discuss a possible mechanism for generating such unwanted spontaneous reversals in a bistable system.

1.4 Stochasticity in molecular networks of developmental processes

Genetic variation describes the presence of genetic differences among individuals. Those difference can arise due to, e.g. genetic mutations, homologous recombination, and immigration. Genetic variation permits flexibility and survival of a population when the environment is dynamic in time. Subsequently, this form of genetic variation is often considered as an advantage.

The variability observed among individuals cannot be fully explained by genetic variation. Even two genetically identical individuals in the same environmental conditions show variability. These differences can be explained by stochastic variation on molecular levels. Stochastic variation was first shown in gene expression in *E. coli*, in which a synthetic represillator showed considerable cell-to-cell variability in both the amplitude and period of the oscillations due to fluctuation in the expression levels of the components (Elowitz and Leibler 2000). Much of the stochasticity in gene expression is caused by transcriptional bursts, meaning that in individual cells the level of transcription is irregular, with bursts of activity and long periods of inactivity (Golding et al. 2005; Raj et al. 2006). Not only gene expression is inherently stochastic, proteins also show translational bursts (Ozbudak et al. 2002). The impact of the stochastic fluctuations is dependent on the copy numbers of molecules involved. For example, fluctuations are typically small when the copy number of molecules is large.

Stochasticity leads to fluctuations in the components of genetic regulatory

networks that perturb the function of such networks, as shown in the synthetic repressilator in *E. coli*, in which the gene expression shows noisy, variable behavior in the period and amplitude of the oscillations (Elowitz and Leibler 2000). In some cases, the stochasticity is exploited in order to perform a specific developmental program in a stochastic manner, for example in the stochastic choice of cell fate in the generation of photoreceptors in *D. melanogaster* (Wernet et al. 2006). However, most developmental processes must be robust, i.e. unaffected by molecular fluctuations, in order to successfully develop from an egg cell into a full-grown organism (Arias and Hayward 2006). Robustness of developmental processes is often achieved by suppressing stochasticity, and in these cases, stochasticity is seen as an unwanted variable (Raj et al. 2010; Johnston and Desplan 2010).

As stated previously, the genetic regulatory network underlying the dynamics of the terminal selectors responsible for the maintenance of the terminal cell fate, act as genetic switches. The components in these genetic regulatory networks are also subjected to stochastic fluctuations. Moreover, it has been shown that genetic switches can suffer from stochastic transitions between stable states when there are stochastic fluctuations in the levels of their components (Ozbudak et al. 2004; Suel et al. 2006). Hence, if the terminal selectors for bistable genetic switches responsible for long-term cell fate maintenance, it is an important, unanswered question how spontaneous loss of cell fate due to molecular fluctuations is prevented.

1.5 *C. elegans* as a model system

The two subjects of this thesis, developmental timing and maintenance of terminal cell fates, will be addressed using the nematode *Caenorhabditis elegans* as the model system. The nematode *C. elegans*, one of the simplest multi cellular organisms, started to emerge as an important model system after the pioneering work of Sydney Brenner in the 1970s (Brenner 1974; Riddle et al. 1997). It has been used in various fields within biology, ranging from neurobiology to developmental biology. The development of *C. elegans* from a single-celled embryo to an adult containing 959 somatic cells takes only about three days, during which the animals reach a length of approximately 1 mm. During their short life-cycle, the animals go through four larval stages (L1-L4), with each stage ending with the ecdysis, and the renewal and shedding of the skin (cuticle) (Fig. 2). The short life-cycle is one of the desirable traits that makes *C. elegans* easily amendable for experimentation.

Due to the transparency of the eggshell and cuticle, the development of cells can be tracked in time, leading to the establishment of the entire cell lineage from a single-celled embryo to adult animals (Kimble and Hirsh 1979; Sulston and Horvitz

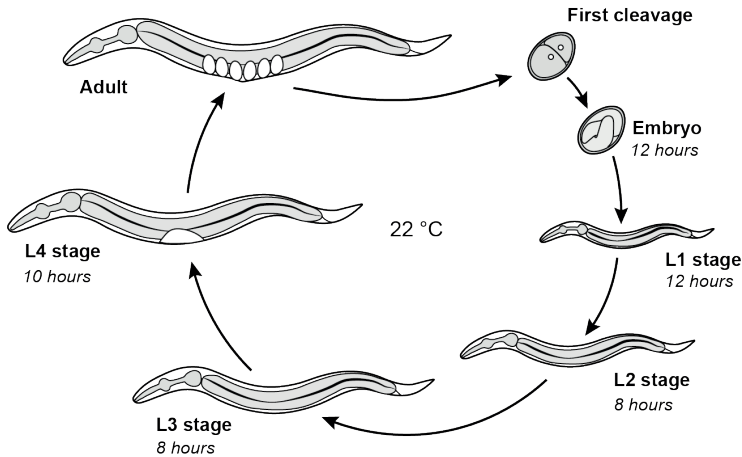


Figure 2: Schematic of *C. elegans* life cycle at 22 °C. *C. elegans* hermaphrodite completes its life cycle in ~3 days at 22°C. When the egg is fertilized, it takes ~12 hours until it completed its embryonic stage. After hatching, animals go through four larval stages, L1-L4. The time given underneath the larval stages represents the approximate duration of each larval stage (Altun 2009).

1977; Sulston et al. 1983). These cell lineages revealed that every animal developed in the same, invariant manner, with almost no difference between the type or number of somatic cells. The transparency of the cuticle also allows for fluorescent microscopy in intact animals. Moreover, recent advantages in gene editing with the CRISPR/Cas9 method (Friedland et al. 2013), made endogenously tagging genes with fluorophores or other tags possible, allowing for quantitative measurements on gene expression and protein levels. Additionally, other quantitative experimental techniques, such as single molecule FISH (smFISH) can be used to quantify gene expression levels. With smFISH, single mRNA molecules can be visualized in fixed animals by the binding of multiple probes that are targeted to the region of the same mRNA, in which the probes have a fluorophore attached (Raj et al. 2008). The smFISH method is widely-used throughout the chapters of this thesis. In **Chapter 2**, I will describe the analysis software package that we designed to aid with smFISH experiments specifically aimed at large *C. elegans* data sets.

In the next sections, I will focus on the maintenance of the terminal cell fate in *C. elegans*, specifically in the neurons. The neurons are not renewed in the animals, and are expected to maintain their specialized functions during the entire lifetime (~2 weeks). Also, the terminal selector genes responsible for the terminal cell fate of neurons are well-studied, making them a suitable subject for studying mechanisms

underlying the maintenance of terminal cell fate. In the last section, I will discuss the potential role of gene expression oscillators in regulation of the worm's molting cycle, which is suggested to act as a developmental clock. Whereas it is clear that these gene expression oscillations play a role during the molt, it is an open question whether they are involved in timing of development outside of the molt.

1.6 Terminal cell fate maintenance of ASE neurons

An adult *C. elegans* hermaphrodite has a nervous system, consisting out of 302 neurons and 56 glial cells (White et al. 1986; Ward et al. 1975; Sulston and Horvitz 1977; Sulston 1983), which is by far the most complex organ of the animal. The nervous system can be divided into two independent nervous systems: 20 neurons are part of the pharyngeal nervous system located in the pharynx, and the remaining 282 neurons make out the somatic nervous system located in the head and tail ganglia, and along the ventral nerve cord. Most of the neurons develop during embryogenesis, with only 80 neurons developing post-embryonically. When the neurons are fully developed, they all have specialized functions, such as different sensory neurons responsible for detecting various soluble and volatile compounds, touch stimuli and sensing temperature. Other groups of neurons are interneurons, involved in information processing of signals of other neurons. And motor neurons, responsible for controlling muscle movement.

The terminal differentiation of those neuronal cells into specialized neurons in *C. elegans* has been well-studied, and lead to the discovery that the terminal cell fate of many neurons is controlled by only a few terminal selector genes in each individual neuron (Hobert 2008, 2016). One of the most studied terminal selector genes in *C. elegans*, is the zinc-finger transcription factor *che-1* expressed in the two ASE chemosensory neurons, ASER and ASEL, in *C. elegans* (Fig 3A) (Chang, Johnston, and Hobert 2003; Uchida et al. 2003). The ASE neurons sense small soluble molecules in the external environment, such as sodium chloride. The ASER and ASEL ordinate from two separate lineage branches. Nevertheless, they are bilaterally symmetrical and structurally similar. However, ASER and ASEL are functionally different, e.g. ASER responds to a decrease in Cl⁻, while ASER responds to an increase in Na⁺ (Ortiz et al. 2009; Smith et al. 2013). CHE-1 is responsible for the induction of the target genes in both ASE neurons, by binding 12-base-pair ASE motifs in the promoter region of the target genes (Etchberger et al. 2007). In addition, CHE-1 induces its own expression, a key feature of a terminal selector gene. During the embryonic development, *che-1* expression is induced by the Tailless/TLX transcription factor NHR-67 (Sarin et al. 2009). After the transient induction, *che-1* expression has to be maintained throughout the lifetime of the animal, since it is required for the continuous induction of its target genes (Etchberger et al. 2009).

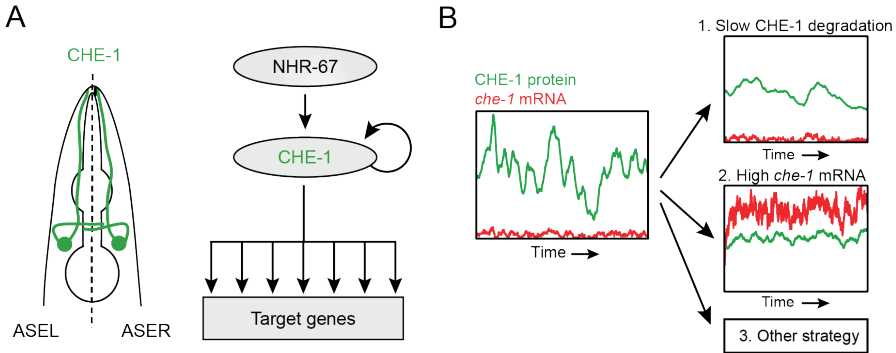


Figure 3. Maintenance of ASE terminal cell fate by the CHE-1 switch and its possible strategies. (A) Schematic of the head of *C. elegans* with the ASER and ASEL highlighted in green where CHE-1 is expressed. The schematic of the CHE-1 switch shows the transient induction of NHR-67 and the maintenance of *che-1* expression by positive autoregulation, while CHE-1 induces hundreds of target genes. (B) Stochastic model of *che-1* mRNA and CHE-1 protein, showing potential strategies to overcome stochastic fluctuations. Well-known strategies are increasing the lifetime of components (e.g. slow CHE-1 protein degradation) or increase molecular copy numbers (e.g. high *che-1* mRNA copy number). We identified a different and much more powerful mechanism in **Chapter 3**.

The CHE-1 terminal selector network is one of the simplest known, with only one positive autoregulatory feedback loop, resembling the classic architecture of a genetic switch. As I described in more detail in Section 1.2, positive autoregulatory networks exhibit two steady states. In the first state, CHE-1 has low/no expression. When CHE-1 is not expressed, the target genes are no longer induced, subsequently the ASE function is lost. In the second state, CHE-1 is expressed in high levels. After the induction of NHR-67, CHE-1 has to maintain its own expression by the positive autoregulation during the entire lifetime of an animal. Stochastic fluctuations in CHE-1 could potentially lead to an unwanted transition between the two states. A straightforward solution would be to have either high copy numbers or slow degradation rates of the CHE-1 protein (Fig. 3B). However, additional mechanisms could be in place to stabilize CHE-1 levels in order to prevent spontaneous switch off. In **Chapter 3**, we identify a novel mechanism for life-long maintenance of ASE neuron fate in *C. elegans* by the CHE-1 genetic switch, that ensure extreme stability of the switch even for low copy numbers and short life-times of *che-1* mRNA and protein molecules.

1.7 Oscillating genes in the pharynx

Each larval stage ends with the molt, during which the animal is in a quiescent state, called lethargus. At the end of the molt, the skin of the animal, called the cuticle,

is renewed and shed (Monsalve and Frand 2012; Lazetic and Fay 2017). Prior to the molt comes the intermolt, during which the animal is actively feeding. Those recurring events during the development of a larva were shown by a study of Johnstone et al. to coincide with rhythmic expression of genes known to be part of the composition of the cuticle (Johnstone 1994). More recent studies showed that even approximately 1/5th of the larval transcriptome in *C. elegans* consists out of genes that are rhythmically expressed throughout larval development (Hendriks et al. 2014; Meeuse et al. 2020; Kim, Grun, and van Oudenaarden 2013). The oscillatory period of those genes coincides with the larval stages, and when temperature decreases the development slows down as well as the oscillating period (Grun et al. 2014; Hendriks et al. 2014). Those observations suggest a link between the rhythmically expressed genes and development, where the genes act as a developmental clock.

Among the genes that are rhythmically expressed during the larval development, the genes involved in the molting cycle are highly enriched, such as cuticular collagen components and proteases required for modification of the cuticle (Hendriks et al. 2014; Meeuse et al. 2020). However, a large number of genes has a phase that peaks outside the molt period and show no apparent link with molting. It remains elusive what the biological function is of those genes. A subset of genes peaking outside the molt and showing no apparent link with molting are expressed in the pharynx (Meeuse et al. 2020), a neuromuscular organ used for pumping food (Fig 4A-B). The bulk of the pharynx is made up out of muscles cells that are divided into eight sections, and marginal cells that are divided into three sections, along the longitudinal axis. Other cell types and structures found in the pharynx include five gland cells, twenty neurons, six epithelium cells and a cuticle lining the pharynx lumen. During larval development, the pharynx grows and the cuticle of the pharynx is renewed and shed, similar to the body cuticle.

So far, the presence and function of the subset of rhythmically expressed genes in the pharynx has not been studied. In **Chapter 5**, I will focus on the function of those oscillating genes in relation to the pharynx growth, which can be seen as a developmental event occurring periodically in time during the larval development. When development is arrested upon starvation, the growth of the pharynx is expected to be interrupted. By combining such developmental arrests with measurements of pharynx growth and oscillatory gene expression, we examine whether the oscillating gene expression is functionally linked with the growth of the pharynx.

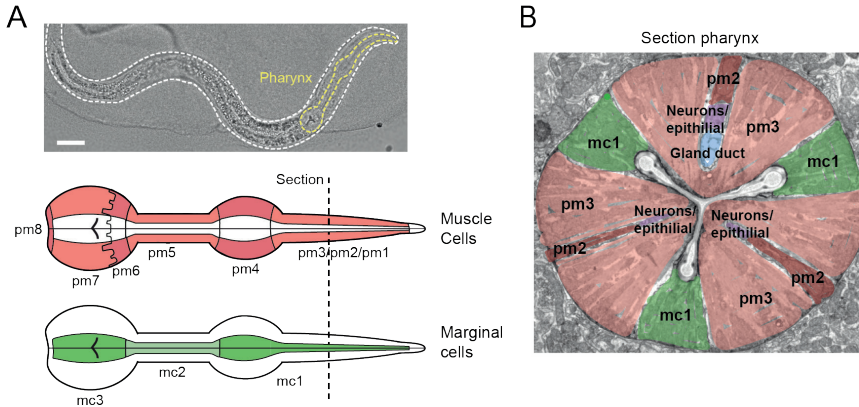


Figure 4: Pharynx anatomy of *C. elegans*. (A) Microscopy image of a *C. elegans* L2 larva, highlighting the location of the pharynx in yellow. White lines: body outline. Scale bar: 20 μ m. Bottom panels show schematic of the location of pharyngeal muscle and marginal cells in the pharynx. The pharynx is composed of eight muscle cells (pm1-8, red), separated by three marginal cells (mc1-3, green), along the longitudinal axis. (B) The muscle and marginal cells are arranged with a three-fold rotational symmetry. And embeds several other cell types, including the gland duct, neurons and epithelial cells (Altun 2020).

1.8 Scope of this thesis

The contents of this thesis are divided in the following chapters:

Chapter 2 describes the software package we wrote for analysing smFISH data. The highly precise measurements of gene expression enabled by smFISH were crucial for quantification of terminal selector genes (**Chapter 3**) and oscillatory gene expression (**Chapters 4** and **5**). The package we developed is specifically aimed at the type of data that is produced with smFISH experiments on *C. elegans*. In contrast to existing smFISH analysis packages that focus on single cells, *C. elegans* data requires many more smFISH microscopy images and the ability to annotate cells within the animal. We optimized the analysis workflow for *C. elegans* data to reduce the time spend on analysis. In addition, we added a tool to the package that allows for easy annotation of the developmental stage of the specific individual.

Chapter 3 explores the underlying mechanisms behind the maintenance of terminal cell fate by the genetic cell fate switch involving the transcription factor CHE-1, required for the induction and maintenance of the terminal cell fate of the ASE neurons. We measured the parameters of the CHE-1 switch. Next, we combined these measured parameters with stochastic simulations to uncover a mechanism that explains life-long maintenance of ASE neuron fate. In this mechanism,

continued *che-1* expression is ensured by preferential binding of CHE-1 to the *che-1* promoter. We validated this mechanism by transient depletion of CHE-1 levels in the ASE neurons. Additionally, we identified a 130 bp fragment in the *che-1* promoter region responsible for the stable *che-1* expression. Moreover, we also identified an *Otx*-related homeodomain binding site within the fragment involved in the stabilization of *che-1* expression to prevent spontaneous loss of CHE-1 expression and subsequently ASE function.

In **Chapter 4**, we set out to validate the existence of two clusters found with a novel percolation clustering algorithm in RNA-sequencing data on young adult animals (Werner et al. 2020). We also searched for a possible origin underlying the correlated variation in expression between genes in these clusters. We quantified the mRNA levels of selected genes from the clusters in animals from similar age as in the RNA-sequencing data, showing that indeed at least one of the clusters consists of genes that are correlated in expression, validating one cluster. The analysis showed that the validated cluster was most likely the result of oscillations in gene expression related to development. Correlation between the selected genes of both clusters increased when starvation was introduced, by inducing in one cluster a temporal expression pattern resembling that of the other cluster.

Chapter 5 studies the gene cluster we validated in Chapter 4, with genes specifically expressed in the pharynx, focusing on its relation with the molting cycle oscillations and pharynx growth. By following the growth of the pharynx in time, we showed that the growth of the pharynx has a cyclical character. The oscillating expression of the selected genes appeared to be coordinated with the growth dynamics of the pharynx in development. By arresting the development of the animals through starvation, we found that in general, if pharynx growth halted, so did oscillatory gene expression in the pharynx, suggesting that gene expression oscillations and pharynx growth were functionally coupled.

1.9 References

- Abrieu, A., M. Doree, and D. Fisher. 2001. 'The interplay between cyclin-B-Cdc2 kinase (MPF) and MAP kinase during maturation of oocytes,' *J Cell Sci*, 114: 257-67.
- Alon, U. 2007. 'Network motifs: theory and experimental approaches,' *Nat Rev Genet*, 8: 450-61.
- Altun, Z.F. & Hall, D.H. 2009. 'Introduction. In WormAtlas. '
- Altun, Z.F., Crocker, C. and Hall, D.H. 2020. "In WormAtlas." In, SW-Worm Viewer, Slice No. (30). .
- Arias, A. M., and P. Hayward. 2006. 'Filtering transcriptional noise during development: concepts and mechanisms,' *Nat Rev Genet*, 7: 34-44.
- Brandman, O., J. E. Ferrell, Jr., R. Li, and T. Meyer. 2005. 'Interlinked fast and slow positive feedback loops drive reliable cell decisions,' *Science*, 310: 496-8.
- Brenner, S. 1974. 'The genetics of *Caenorhabditis elegans*,' *Genetics*, 77: 71-94.
- Chang, S., R. J. Johnston, Jr., and O. Hobert. 2003. 'A transcriptional regulatory cascade that controls left/right asymmetry in chemosensory neurons of *C. elegans*,' *Genes Dev*, 17: 2123-37.
- Darwich, A. S., U. Aslam, D. M. Ashcroft, and A. Rostami-Hodjegan. 2014. 'Meta-analysis of the turnover of intestinal epithelia in preclinical animal species and humans,' *Drug Metab Dispos*, 42: 2016-22.
- Duggan, A., C. Ma, and M. Chalfie. 1998. 'Regulation of touch receptor differentiation by the *Caenorhabditis elegans* mec-3 and unc-86 genes,' *Development*, 125: 4107-19.
- Ebisuya, M., and J. Briscoe. 2018. 'What does time mean in development?,' *Development*, 145.
- Eissing, T., H. Conzelmann, E. D. Gilles, F. Allgower, E. Bullinger, and P. Scheurich. 2004. 'Bistability analyses of a caspase activation model for receptor-induced apoptosis,' *J Biol Chem*, 279: 36892-7.
- Elowitz, M. B., and S. Leibler. 2000. 'A synthetic oscillatory network of transcriptional regulators,' *Nature*, 403: 335-8.
- Etchberger, J. F., E. B. Flowers, R. J. Poole, E. Bashllari, and O. Hobert. 2009. 'Cis-regulatory mechanisms of left/right asymmetric neuron-subtype specification in *C. elegans*,' *Development*, 136: 147-60.
- Etchberger, J. F., A. Lorch, M. C. Sleumer, R. Zapf, S. J. Jones, M. A. Marra, R. A. Holt, D. G. Moerman, and O. Hobert. 2007. 'The molecular signature and cis-regulatory architecture of a *C. elegans* gustatory neuron,' *Genes Dev*, 21: 1653-74.
- Ferrell, J. E., Jr. 2002. 'Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability,' *Curr Opin Cell Biol*, 14: 140-8.
- Ferrell, J. E., Jr., and E. M. Machleder. 1998. 'The biochemical basis of an all-or-none cell fate switch in *Xenopus* oocytes,' *Science*, 280: 895-8.
- Ferrell, J. E., Jr., J. R. Pomerening, S. Y. Kim, N. B. Trunnell, W. Xiong, C. Y. Huang, and E. M. Machleder. 2009. 'Simple, realistic models of complex biological processes: positive feedback and bistability in a cell fate switch and a cell cycle oscillator,' *FEBS Lett*, 583: 3999-4005.
- Ferrell, J. E., and W. Xiong. 2001. 'Bistability in cell signaling: How to make continuous processes discontinuous, and reversible processes irreversible,' *Chaos*, 11: 227-36.
- Friedland, A. E., Y. B. Tzur, K. M. Esvelt, M. P. Colaiacovo, G. M. Church, and J. A. Calarco. 2013. 'Heritable genome editing in *C. elegans* via a CRISPR-Cas9 system,' *Nat Methods*, 10: 741-3.
- Gabilondo, H., J. Stratmann, I. Rubio-Ferrera, I. Millan-Crespo, P. Contero-Garcia, S.

- Bahrampour, S. Thor, and J. Benito-Sipos. 2016. 'Neuronal Cell Fate Specification by the Convergence of Different Spatiotemporal Cues on a Common Terminal Selector Cascade,' *PLoS Biol*, 14: e1002450.
- Gardner, T. S., C. R. Cantor, and J. J. Collins. 2000. 'Construction of a genetic toggle switch in *Escherichia coli*,' *Nature*, 403: 339-42.
- Golding, I., J. Paulsson, S. M. Zawilski, and E. C. Cox. 2005. 'Real-time kinetics of gene activity in individual bacteria,' *Cell*, 123: 1025-36.
- Grun, D., M. Kirchner, N. Thierfelder, M. Stoeckius, M. Selbach, and N. Rajewsky. 2014. 'Conservation of mRNA and protein expression during development of *C. elegans*,' *Cell Rep*, 6: 565-77.
- Hendriks, G. J., D. Gaidatzis, F. Aeschmann, and H. Grosshans. 2014. 'Extensive oscillatory gene expression during *C. elegans* larval development,' *Mol Cell*, 53: 380-92.
- Hirota, J., and P. Mombaerts. 2004. 'The LIM-homeodomain protein Lhx2 is required for complete development of mouse olfactory sensory neurons,' *Proc Natl Acad Sci U S A*, 101: 8751-5.
- Hobert, O. 2008. 'Regulatory logic of neuronal diversity: terminal selector genes and selector motifs,' *Proc Natl Acad Sci U S A*, 105: 20067-71.
- Hobert, O. 2016. 'Terminal Selectors of Neuronal Identity,' *Curr Top Dev Biol*, 116: 455-75.
- Hsiao, H. Y., D. Jukam, R. Johnston, and C. Desplan. 2013. 'The neuronal transcription factor erect wing regulates specification and maintenance of *Drosophila* R8 photoreceptor subtypes,' *Dev Biol*, 381: 482-90.
- Johnston, R. J., Jr., and C. Desplan. 2010. 'Stochastic mechanisms of cell fate specification that yield random or robust outcomes,' *Annu Rev Cell Dev Biol*, 26: 689-719.
- Johnstone, I. L. 1994. 'The cuticle of the nematode *Caenorhabditis elegans*: a complex collagen structure,' *Bioessays*, 16: 171-8.
- Jukam, D., and C. Desplan. 2010. 'Binary fate decisions in differentiating neurons,' *Curr Opin Neurobiol*, 20: 6-13.
- Kageyama, R., Y. Niwa, A. Isomura, A. Gonzalez, and Y. Harima. 2012. 'Oscillatory gene expression and somitogenesis,' *Wiley Interdiscip Rev Dev Biol*, 1: 629-41.
- Kim, Dh, D. Grun, and A. van Oudenaarden. 2013. 'Dampening of expression oscillations by synchronous regulation of a microRNA and its target,' *Nat Genet*, 45: 1337-44.
- Kimble, J., and D. Hirsh. 1979. 'The postembryonic cell lineages of the hermaphrodite and male gonads in *Caenorhabditis elegans*,' *Dev Biol*, 70: 396-417.
- Konstantinides, N., K. Kapuralin, C. Fadil, L. Barboza, R. Satija, and C. Desplan. 2018. 'Phenotypic Convergence: Distinct Transcription Factors Regulate Common Terminal Features,' *Cell*, 174: 622-35 e13.
- Lazetic, V., and D. S. Fay. 2017. 'Molting in *C. elegans*,' *Worm*, 6: e1330246.
- Leyva-Diaz, E., and O. Hobert. 2019. 'Transcription factor autoregulation is required for acquisition and maintenance of neuronal identity,' *Development*, 146.
- Meeuse, M. W., Y. P. Hauser, L. J. Morales Moya, G. J. Hendriks, J. Eglinger, G. Bogaarts, C. Tsiairis, and H. Grosshans. 2020. 'Developmental function and state transitions of a gene expression oscillator in *Caenorhabditis elegans*,' *Mol Syst Biol*, 16: e9498.
- Ming, G. L., and H. Song. 2005. 'Adult neurogenesis in the mammalian central nervous system,' *Annu Rev Neurosci*, 28: 223-50.
- Monsalve, G. C., and A. R. Frand. 2012. 'Toward a unified model of developmental timing: A "molting" approach,' *Worm*, 1: 221-30.
- Oates, A. C., L. G. Morelli, and S. Ares. 2012. 'Patterning embryos with oscillations: structure, function and dynamics of the vertebrate segmentation clock,' *Development*, 139: 625-39.

- Ortiz, C. O., S. Faumont, J. Takayama, H. K. Ahmed, A. D. Goldsmith, R. Pocock, K. E. McCormick, H. Kunitomo, Y. Iino, S. Lockery, and O. Hobert. 2009. 'Lateralized gustatory behavior of *C. elegans* is controlled by specific receptor-type guanylyl cyclases', *Curr Biol*, 19: 996-1004.
- Ozbudak, E. M., M. Thattai, I. Kurtser, A. D. Grossman, and A. van Oudenaarden. 2002. 'Regulation of noise in the expression of a single gene', *Nat Genet*, 31: 69-73.
- Ozbudak, E. M., M. Thattai, H. N. Lim, B. I. Shraiman, and A. Van Oudenaarden. 2004. 'Multistability in the lactose utilization network of *Escherichia coli*', *Nature*, 427: 737-40.
- Ptashne, M., and P. Mark. 1992. *A Genetic Switch: Phage [lambda] and Higher Organisms* (Cell Press).
- Raj, A., C. S. Peskin, D. Tranchina, D. Y. Vargas, and S. Tyagi. 2006. 'Stochastic mRNA synthesis in mammalian cells', *PLoS Biol*, 4: e309.
- Raj, A., S. A. Rifkin, E. Andersen, and A. van Oudenaarden. 2010. 'Variability in gene expression underlies incomplete penetrance', *Nature*, 463: 913-8.
- Raj, A., P. van den Bogaard, S. A. Rifkin, A. van Oudenaarden, and S. Tyagi. 2008. 'Imaging individual mRNA molecules using multiple singly labeled probes', *Nat Methods*, 5: 877-9.
- Riddle, D. L., T. Blumenthal, B. J. Meyer, and J. R. Priess. 1997. 'Introduction to *C. elegans*' in nd, D. L. Riddle, T. Blumenthal, B. J. Meyer and J. R. Priess (eds.), *C. elegans II* (Cold Spring Harbor (NY)).
- Sarin, S., C. Antonio, B. Tursun, and O. Hobert. 2009. 'The *C. elegans* Tailless/TLX transcription factor nhr-67 controls neuronal identity and left/right asymmetric fate diversification', *Development*, 136: 2933-44.
- Smith, H. K., L. Luo, D. O'Halloran, D. Guo, X. Y. Huang, A. D. Samuel, and O. Hobert. 2013. 'Defining specificity determinants of cGMP mediated gustatory sensory transduction in *Caenorhabditis elegans*', *Genetics*, 194: 885-901.
- Spalding, K. L., R. D. Bhardwaj, B. A. Buchholz, H. Druid, and J. Frisen. 2005. 'Retrospective birth dating of cells in humans', *Cell*, 122: 133-43.
- Suel, G. M., J. Garcia-Ojalvo, L. M. Liberman, and M. B. Elowitz. 2006. 'An excitable gene regulatory circuit induces transient cellular differentiation', *Nature*, 440: 545-50.
- Sulston, J. E. 1983. 'Neuronal cell lineages in the nematode *Caenorhabditis elegans*', *Cold Spring Harb Symp Quant Biol*, 48 Pt 2: 443-52.
- Sulston, J. E., and H. R. Horvitz. 1977. 'Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans*', *Dev Biol*, 56: 110-56.
- Sulston, J. E., E. Schierenberg, J. G. White, and J. N. Thomson. 1983. 'The embryonic cell lineage of the nematode *Caenorhabditis elegans*', *Dev Biol*, 100: 64-119.
- Toma, K., T. C. Wang, and C. Hanashima. 2016. 'Encoding and decoding time in neural development', *Dev Growth Differ*, 58: 59-72.
- Turek, M., and H. Bringmann. 2014. 'Gene expression changes of *Caenorhabditis elegans* larvae during molting and sleep-like lethargus', *PLoS One*, 9: e113269.
- Tyson, J. J., K. C. Chen, and B. Novak. 2003. 'Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell', *Curr Opin Cell Biol*, 15: 221-31.
- Tyson, J. J., and B. Novak. 2010. 'Functional motifs in biochemical reaction networks', *Annu Rev Phys Chem*, 61: 219-40.
- Uchida, O., H. Nakano, M. Koga, and Y. Ohshima. 2003. 'The *C. elegans* che-1 gene encodes a zinc finger transcription factor required for specification of the ASE chemosensory neurons', *Development*, 130: 1215-24.
- Ward, S., N. Thomson, J. G. White, and S. Brenner. 1975. 'Electron microscopical

- reconstruction of the anterior sensory anatomy of the nematode *Caenorhabditis elegans*.?2UU', *J Comp Neurol*, 160: 313-37.
- Werner, Steffen, W Mathijs Rozemuller, Annabel Ebbing, Anna Alemany, Joleen Traets, Jeroen S. van Zon, Alexander van Oudenaarden, Hendrik C. Korswagen, Greg J. Stephens, and Thomas S. Shimizu. 2020. 'Functional modules from variable genes: Leveraging percolation to analyze noisy, high-dimensional data', *bioRxiv*: 2020.06.10.143743.
- Wernet, M. F., E. O. Mazzoni, A. Celik, D. M. Duncan, I. Duncan, and C. Desplan. 2006. 'Stochastic spineless expression creates the retinal mosaic for colour vision', *Nature*, 440: 174-80.
- White, J. G., E. Southgate, J. N. Thomson, and S. Brenner. 1986. 'The structure of the nervous system of the nematode *Caenorhabditis elegans*', *Philos Trans R Soc Lond B Biol Sci*, 314: 1-340.
- Wolpert, L. 2015. *Principles of development*.
- Xiong, W., and J. E. Ferrell, Jr. 2003. 'A positive-feedback-based bistable 'memory module' that governs a cell fate decision', *Nature*, 426: 460-5.
- Xue, D., Y. Tu, and M. Chalfie. 1993. 'Cooperative interactions between the *Caenorhabditis elegans* homeoproteins UNC-86 and MEC-3', *Science*, 261: 1324-8.
- Yao, G., T. J. Lee, S. Mori, J. R. Nevins, and L. You. 2008. 'A bistable Rb-E2F switch underlies the restriction point', *Nat Cell Biol*, 10: 476-82.

Chapter 2

2

WormFISH: a software package for single molecule FISH analysis of *C. elegans* larval development

This chapter is a part of the following publication:
“Joleen J.H. Traets and Jeroen S. van Zon. WormFISH: a software package for single molecule FISH analysis of *C. elegans* larval development.”
(Manuscript in preparation)

2.1 Abstract

The single molecule fluorescence in situ hybridization (smFISH) method has become an important tool to study transcriptional regulation in the nematode worm *Caenorhabditis elegans*, by visualizing individual mRNA molecules. However, analyzing smFISH data sets obtained from *C. elegans* animals comes with a number of unique challenges, such as the large number of required microscopy images and variation in mRNA detection quality between and within microscopy images. Here, we present a software package, WormFISH, that offers an optimized workflow specifically aimed at analyzing large *C. elegans* smFISH data sets. In addition, WormFISH offers an integrated tool to annotate developmental stage by morphology and anatomical features, allowing a precise developmental age to be assigned to each smFISH measurement. This makes it possible to use smFISH to measure gene expression dynamics over the course of *C. elegans* development.

2.2 Introduction

Changes in gene expression over time are central to cell differentiation and development. Fluorescence in situ hybridization (FISH) is developed to detect DNA and RNA in fixed tissue. It became an important method to visualize gene expression, by using fluorescent labelled DNA binding to complementary mRNA in tissue. The ability to visualize single mRNA molecules with the single molecule FISH (smFISH) method was first demonstrated by Femino et al., by using multiple DNA probes of 50 base pairs with 5 to 6 fluorophores attached targeted to individual actin mRNA molecules (Femino et al. 1998). Subsequently, the smFISH method was adapted to visualize single mRNA molecules in yeast, mammalian cell lines, fruit fly embryos, nematodes and even on whole organoids (Raj et al. 2008; Little, Tikhonov, and Gregor 2013; Omerzu et al. 2019). Most current smFISH methods use multiple (30-100) short oligonucleotides, of approximately 20 base pairs long, each conjugated with a single fluorophore and complementary to a different region of the mRNA molecule of the gene of interest. While a single fluorophore results in a very dim signal, the binding of multiple oligonucleotides on one targeted mRNA yields a sufficiently bright fluorescent spot, representing one single mRNA molecule. The smFISH method makes it possible to unambiguously count the amount of mRNA molecules in individual cells of fixed unicellular or multicellular organisms.

The smFISH method enables detailed visualization of spatial gene expression patterns in tissues and embryos, providing important information on biological function. Moreover, it makes it possible to quantify gene expression with single mRNA precision and characterize mRNA localization with subcellular resolution. In the past decades, the method has been applied successfully in numerous studies examining stochastic gene expression, resulting in heterogeneity among otherwise identical cells (Chong et al. 2014; Little, Tikhonov, and Gregor 2013; Raj and van Oudenaarden 2008; Sanchez and Golding 2013; Raj et al. 2006). More recent findings in tissue samples have revealed important insight into the mechanism and biological function of heterogeneity, for example, with the zonation of the mammalian liver for the division of labor between liver cells (Halpern et al. 2017), or the mechanism of suppressing noise in gene expression via auto-depletion (Hansen et al. 2018). Here, the precise quantification of expression levels by smFISH is essential to distinguish stochastic gene expression from experimental noise.

After smFISH staining, mRNA molecules are visible as diffraction-limited spots with a Gaussian fluorescence intensity profile (Fig. 1A-C). Such smFISH spots can be counted by hand if only a few are present. However, counting highly or broadly expressed genes by hand is no longer feasible. A number of software packages have been developed for (semi)-automated detection and counting of smFISH spots (Raj

et al. 2008; Mueller et al. 2013; Tsanov et al. 2016; Pharris et al. 2017; Rifkin 2011; Stoeger et al. 2015). Currently, the most used software package for semi-automated smFISH spot counting is FISH-quant (Mueller et al. 2013), aimed at a broad spectrum of model systems. In general, most packages for counting smFISH spots provide a streamlined image analysis pipeline with semi-automated identification of the fluorescent spots. In addition, some packages provide automated cell recognition based on cell nuclei staining, to obtain smFISH spot counts with cellular resolution.

The smFISH method has become an important tool to study transcriptional regulation in the nematode worm *Caenorhabditis elegans* (Ji and van Oudenaarden 2012). In the past decade, it has been successfully applied in numerous studies in *C. elegans* (Huelsz-Prince and van Zon 2017; Frokjaer-Jensen et al. 2016; Lee et al. 2016; Middelkoop et al. 2012; Harterink et al. 2011; Topalidou, van Oudenaarden, and Chalfie 2011). However, existing software packages are not well-suited for analyzing smFISH data sets obtained for *C. elegans* animals, as these come with a number of unique challenges. First, most existing packages are optimized for the analysis of smFISH data obtained in cell culture, where a relatively small number of microscopy image stacks each contain many identical cells. In contrast, smFISH data for *C. elegans* typically consist of a large number of microscopy image stacks, that each contain a single animal in which a number of qualitatively distinct cells are analyzed, thereby requiring a significantly different workflow. Second, the quality of smFISH spots varies among individual animals within a single experimental batch, meaning that batch automation of smFISH analysis, which is an integral part of existing analysis packages, is not easily possible. Third, due to the comparatively large size of *C. elegans* animals and the resulting light scattering within the animal's tissue, the quality of smFISH spots varies between image slices acquired at different depths within the animal. This makes it difficult to use a single criterion, e.g. a fluorescence intensity threshold, to separate smFISH spots from background noise, as can be done for smFISH data obtained in cell culture. Therefore, we developed a software package, WormFISH, optimized for *C. elegans* smFISH analysis that specifically addresses these challenges.

In general, because samples must be fixed for staining, smFISH experiments are not naturally suited for studying gene expression dynamics. Here, a unique advantage is provided by the highly invariant nature of *C. elegans* development (Sulston and Horvitz 1977). As a result, the animal's anatomy, such as its body length, the length of the gonadal arms and the morphology of its reproductive organs, can be used as a measure of time. For that reason, we integrated our smFISH software package with manual annotation tools to extract these features from microscopy images acquired in parallel with smFISH images, allowing a precise developmental age to be assigned to each smFISH measurement.

WormFISH uses a fluorescent spot identification algorithm similar to Raj et al. (Raj et al. 2008), for detecting fluorescent spots in three dimensions by convolution and filtering smFISH images. In addition, WormFISH provides tools to define multiple regions of interests to quantify gene expression specifically within individual cells or larger anatomical structures, and offers the possibility to set spot selection criteria by hand for each different region of interest. The user interface is optimized for rapid switching both between annotation at the scale of the entire animal and that of individual cells, and between individual animals within one experiment. We show that WormFISH identifies smFISH spots (mature mRNAs) with a precision similar to FISH-quant. Finally, we present a number of examples that demonstrate the power of WormFISH in characterizing gene expression dynamics during *C. elegans* larval development.

2.3 Results

2.3.1 Overview of smFISH protocol

An essential first step for quantification for smFISH imaging data is obtaining high quality fluorescent smFISH spots. Detailed smFISH protocols for *C. elegans* smFISH are found in the study of Ji et al. (Ji and van Oudenaarden 2012). Here, we provide a short overview of important considerations for smFISH and briefly outline the protocol we followed for smFISH staining.

First, we need to create a probe set for the mRNA transcript of the gene of interest, aimed at getting high quality smFISH data. We designed our oligonucleotide probes using the Stellaris probe designer from Biosearch (Biosearch Technologies, Inc., Petaluma, CA, USA), which yields a list of probe sequences with optimized CG contents. Subsequently, we used BLAST to ensure that the oligonucleotides have low affinity for off-target sites on other transcripts in *C. elegans*. In our hands, 20 oligonucleotides specific for one mRNA transcript still yielded a reliable signal (data not shown). However, 25-48 oligonucleotides targeted at one transcript are recommended. The smFISH probe libraries were coupled with fluorophores in-house, following a previously published protocol (Lyubimova et al. 2013). The fluorophores Cy5 and Alexa594 are commonly used, and have a good signal-to-noise ratio. In general, fluorophores with similar long wavelengths are recommended due to increased autofluorescence in shorter wavelengths (Ji and van Oudenaarden 2012). The fluorophore Cy5 yielded a better signal to noise ratio compared to Alexa594, due to its lower autofluorescence levels (Fig 1A-C). Hence, in general, analysis of Alexa594 smFISH spots is more challenging compared to Cy5.

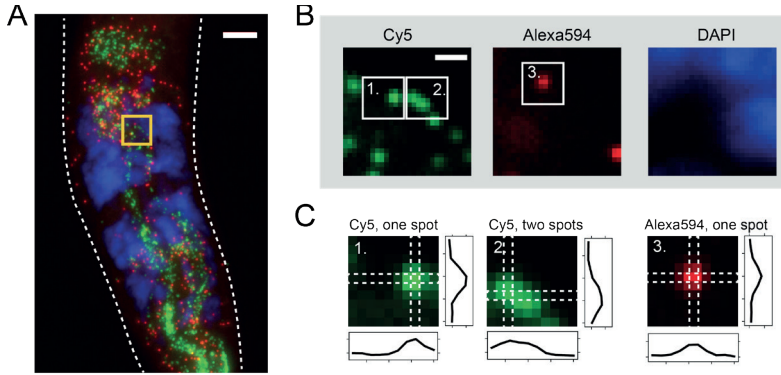


Figure 1. Comparison of Cy5 or Alexa594 smFISH spots. (A) Example of a wide-field fluorescence microscopy image of the head of a *C. elegans* animal at the L1 larval stage, showing smFISH spots labelled with Cy5 (*marg-1* mRNAs, green) and Alexa594 (*nhr-7* mRNAs, red). DAPI staining is used to visualize cell nuclei (blue). Dashed white lines represent the outline of the animal's head. Image is a maximum projection. Scale bar: 5 μm . (B) Enlarged view of the region of interest in panel (A), showing the three channels separately: Cy5, Alexa594 and DAPI. Each image is a single optical slice. Scale bar: 1 μm . (C) Marked spots from right top panel, are analysed on intensity in x and y direction, showing the difficulty of separating two fluorescent spots (2) and the quality difference between both fluorophores (1 vs. 3). White lines represent the column or row of which the pixel intensities are shown in the plots on the x and y axes.

We used the following protocol to perform smFISH staining with coupled probes (see Ji et al. and Methods for a detailed protocol): first, the animals are fixed with formaldehyde (4%) for 40 minutes. For animals older than the L3 staged animals, smFISH spot quality decreased due to higher autofluorescence within the animal's body. After the fixation, the animals were washed with 1X PBS and kept in ethanol for at least overnight. The next day, the fixed animals were hybridized with the probes, and then incubated overnight at 37°C. On the day of imaging, the fixed animals are washed at least 3x with washing solution and DAPI staining was added. After the staining, the fixed animals are stored in 2X SSC until imaging. For imaging, the fixed animals are transferred into the GLOX buffer (with catalase and glucose oxidase) to prevent bleaching while imaging, and added as a droplet onto a cover glass with a round cover glass subsequently placed on top of the droplet. Older animals (>L3 stage) need to be squished as much as possible in-between the glasses, via gentle pressure and removal of excess GLOX buffer with tissue paper. After adding the GLOX buffer, images of the fixed animals are taken immediately with a wide-field fluorescence microscope.

Because of the small size of mRNA molecules, upon smFISH staining they appear as diffraction-limited spots with a Gaussian spot intensity profile (Fig 1B-

C). The fluorescent spots need to be distinguished from the background of intrinsic autofluorescence of the animal's tissue and the fluorescence signal of aspecifically bound smFISH probes. The concentration of the probes can be tuned for an optimal result. High probe concentrations lead to high background fluorescence of aspecifically bound probes, while a too low probe concentration results in low smFISH spot intensities, making spots harder to separate from the background. In some cases, mRNA transcripts can be located in close proximity to each other, resulting in overlap of fluorescent spots. Spots that overlap are in general harder to distinguish from each other. Overall, a successful smFISH spot identification algorithm should be able to both distinguish spots from background noise and recognize partially overlapping spots.

2.3.2 Workflow for identifying and counting smFISH spots

The workflow for identifying and counting fluorescent spots can be divided into two parts: first, the software requires user annotation of smFISH images to determine which part of the smFISH image has to be analyzed. Secondly, the spot identifying algorithm is applied on the parts of the smFISH image that has to be analyzed. During the spot identifying algorithm, the smFISH images are convolved and converted to optimize the spot identification in the images. Next, the threshold to separate the fluorescent spots from the background noise needs to be set by hand with aid of several parameters.

Most existing software packages are not well-suited for analyzing smFISH data sets collected for *C. elegans* animals. In general, smFISH experiments performed in cell culture give rise to smFISH image stacks in which many identical cells must be analyzed. In contrast, for *C. elegans*, each smFISH image stack typically contains a single animal in which potentially only a single cell must be analyzed. Hence, maintaining a sufficiently large sample size typically requires analysis of many smFISH image stacks. Therefore, we optimized our software to easily switch between data for individual animals during the entire analysis process, to facilitate comparison between animals. The analysis process starts with identifying all smFISH image stacks to be analyzed within one batch or experiment. Next, points of interest (POIs), corresponding to individual cells or other anatomical structures, are specified in each image stack by the user (Fig. 2A). The POIs are used to generate a zoomed-in subsection of the image stack, allowing for fast annotation of the corresponding region of interest (ROI). The ROIs are user-generated polygons that typically mark the cell body and are based on cell or tissue identification, e.g. using fluorescent reporters as cell markers, or based on the location/appearance of nuclei as visualized by DAPI staining and/or using the localization pattern of smFISH probes with known cell-specific expression patterns. After the ROIs are specified,

the algorithm for identification of the fluorescent spots starts, in which the user needs to define the threshold in order to separate the fluorescent spots from the noisy background.

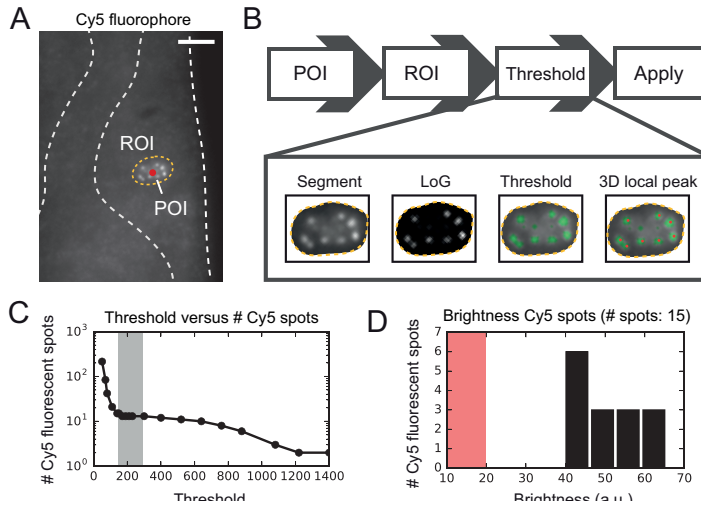


Figure 2. Workflow for identifying smFISH spots. (A) Maximum projection of a Cy5 smFISH image stack (*gcy-14* mRNA, ASEL neuron), with the corresponding region of interest (ROI) and point of interest (POI) shown in orange and red respectively. (B) Scheme of the workflow of spot identification, with example images highlighting the different stages of the threshold process. First the POI and ROI are manually specified. Next, the threshold to distinguish fluorescent spots from the noisy background is set manually with aid of two features shown in (C) and (D). (C) Threshold versus the number of counted Cy5 fluorescent spots, quantified from the example image in (A). The grey inset highlights the plateau on which the choice for the user-determined threshold is based. (D) Brightness histogram of the quantified fluorescent spots of the Cy5 fluorophore (Threshold: 180). The red inset represents the background intensity.

The algorithm for identification of the fluorescent spots is based on previously published algorithms for spot detection (Raj et al. 2008; Ji and van Oudenaarden 2012). During the thresholding process, the ROIs are first segmented from the full smFISH image stacks (Fig. 2A-B). Thereafter, the images are sharpened by convolving the images first by a Gaussian followed by a Laplacian, also called a Laplacian of Gaussian (LoG) filter. The width of the Gaussian filter resembles the size of a typical spot (~5 pixels for our data). This filter smoothens out the background noise and amplifies the contrast between the background and the fluorescent spots. If a threshold would be applied to the image at this stage, overlapping fluorescent

spots might be counted as a single spot. To this problem, we added an extra step to the algorithm: we detangle possible overlapping spots, by using the 3D local peak finding algorithm from the *scikit-image* Python package (van der Walt et al. 2014), which find local maxima in 3D volumes of varying signal intensity. To minimize the time required for this step, the local peak algorithm is only applied to a subset of putative smFISH spots. Specifically, we select for this analysis step 3D volumes of connected pixels with signal intensity higher than the threshold if their volume is larger than that of a typical fluorescent spot (between 10-40 pixels).

After the images have been sharpened and processed, the threshold needs to be set for each of the smFISH images. smFISH images of older animals (>L3 stage) have increased background noise in image slices further away from the microscope objective, due to increase of light scattering by the intervening tissue. Hence, we added the option to set the threshold for each ROI within a smFISH image separately. For selection of the best threshold, we calculate the number of detected spots in each ROI for a range of thresholds (Fig. 2C). In case of high quality smFISH data, the resulting threshold curve shows a transient plateau corresponding to the optimal threshold, where the threshold reliably differentiates between the fluorescent spots and the background noise. smFISH images from the Alexa594 fluorophore or in older animals (>L3 stage) are in general of lower quality, and there the plateau is hardly visible or not present (Fig. S1A-B). In these cases, automatic or semi-automatic thresholding would not work. Instead, we use custom thresholding with aid of three features to guide the user in choosing the best threshold. First, the identified spots in the smFISH image can be made visible during the process, to aid checking the reliability of the identified fluorescent spot at the current set threshold by eye. Next, the threshold range in the threshold curve can be set to custom limits, allowing to zoom in on a specific range of thresholds to search for the plateau. Finally, the user can check if the maximum brightness intensity of the fluorescent spots does not show a too broad range of intensities as an indication for false positives (Fig. 2D). In our hands, the combination of these three features suffices to reliably set a custom threshold for Cy5 and Alexa594 smFISH data of sufficient quality.

2.3.3 Annotation of developmental age in fixed animals

In general, smFISH is not well-suited to study gene expression, as smFISH experiments require fixation of the biological samples. However, the highly invariant nature of *C. elegans* development, makes it possible to use its anatomy as a measure of time (Sulston and Horvitz 1977). To determine the developmental age of individual animals, WormFISH allows for annotation of different anatomical features that correlate strongly with age. Based on previous studies, we included annotation

of five different features to be used for determining the developmental age: 1) body length, 2) gonad arm formation, 2) vulva morphology, 4) spermatogenesis, and 5) expression level of dynamically varying fluorescent reporters (Fig 3A).

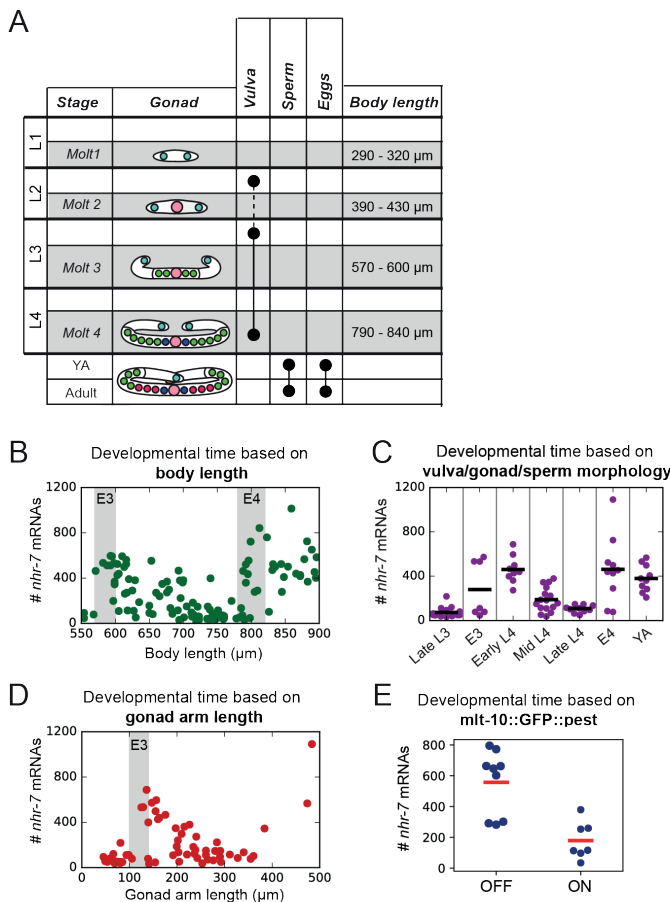


Figure 3. Determining the developmental stage of animals based on features other than body length. (A) Schematic overview of indicators that can be used to aid developmental staging of larval stages in smFISH experiments. L1-4 stages are divided up into molts and intermolts. Morphology information that can be used for specifying developmental stage is shown with lines between black dots. The intermittent line means that it could be informative with addition of markers. Schematic gonad represents the different stages during molts. Yellow: mitotic nuclei, pink: vulva, red: oogenesis, green: meiotic nuclei, blue: sperm, light-blue: DTC. The addition of the vulva morphology offers the possibility to divided the L4 stage up into 10 substages (Mok, Sternberg, and Inoue 2015). Sperm and eggs give information on the adult or young adult stage. The optional *mlt-10::GFP* gives information if animals are in lethargus/molting. (B) Expression levels of the oscillating gene *nhr-7* during developmental time measured with the body length of the individual animals. The approximate molting times of M3 and M4 are represented by the grey insets. Expression levels of *nhr-7* show high variability between the two ecdysis. Approximate body length at ecdysis based on

data from Gritti et al. (Gritti et al. 2016). (C) We used vulva/gonad/sperm morphology to determine developmental stages, showing less variability in expression levels of *nhr-7* compared to the body length. Each substage consists out of the following vulva/gonad/sperm morphologies: late L3: <L4.0, M3: L4.0, early L4: L4.1-3, mid L4: L4.4-5, late L4: L4.6-8, M4: L4.9, YA: sperm. (D) Expression levels of *nhr-7* versus the gonad arm length, with the approximate molting time of M3 represented by the grey insets. The gonad arm length shows similar results on the temporal expression pattern of *nhr-7* as the body length. (E) Expression of *nhr-7* when *mlt-10::GFP::pest* is either high or low, representing the lethargus when its high. The expression of *nhr-7* is low during the lethargus as expected from the previous observations.

The body length increases with a fixed rate that is larval stage-dependent, with growth reducing slightly before molts (Gritti et al. 2016; Uppaluri and Brangwynne 2015; Byerly, Cassada, and Russell 1976), making body length a good indicator of developmental stage. For example, when quantifying the expression levels of the oscillating gene *nhr-7*, a transcription factor expressed exclusively in the pharynx region, the body length provides insight into the time dynamics of the oscillations during the L3 and L4 stage (Fig. 3B). Moreover, from data analysis on RNA-sequencing data set on the entire *C. elegans* larval development of Meeuse et al. (Meeuse et al. 2020), *nhr-7* was predicted to oscillate during larval development. Indeed, we showed that *nhr-7* oscillates during the L3 and L4 stages, showing that body length can provide sufficient information on temporal expression. However, some body lengths show high variability in expression levels of *nhr-7*. Moreover, especially in older animals (>L3 stage), the measured body length is likely less reliable due to deformation of these (larger) animals during imaging.

For that reason, other features than the body length of the animal's anatomy offer more precise determination of the developmental stage. One feature that could be used instead is the development of the gonad during the late L2-YA (Kimble and Hirsh 1979; Kimble and White 1981; Killian and Hubbard 2005; Gritti et al. 2016). The changing gonad formation can be used to determine stages between late L2-YA. To quantify the changing gonad, the gonad arm length or the formation itself can be used as a representation of gonad development. Studying the vulva morphology offers the possibility to divide the L4 stage up into 10 defined stages (Mok, Sternberg, and Inoue 2015). Spermatogenesis morphology can help to differentiate between late L4 stages and young adult animals, when the vulva morphology and gonad arm formation does not change anymore. Finally, another feature that could help with determining the developmental stage, are the expression levels of dynamically varying fluorescent reporters, such as the *mlt-10::GFP::pest* reporter. The reporter can be used as an indicator for when the animals are in lethargus during each molt (Frandsen, Russel, and Ruvkun 2005).

When using morphological features, in this case combining vulva, gonad and sperm morphology to determine developmental stages, we find less variability in

nhr-7 mRNAs (Fig. 3C), suggesting that using this approach improves the accuracy of determining developmental age. Instead of using the formation of the gonad as an indication for developmental stage, we measured the gonad arm length (anterior gonad arm) to check if this approach would provide similar results. Gonad length showed similar variability as the body length in *nhr-7* expression levels for the length measurement (Fig. 3D). Last, the *mlt-10::GFP::pest* expression, as indication for the lethargus confirmed that the *nhr-7* expression is lowest during the lethargus and highest during the intermolt (Fig. 3E). Taken together, the four approaches all provide useful information on the oscillating gene expression, where the developmental time based on the morphology of the vulva/gonad/sperm revealed the most detailed information on the temporal expression pattern of *nhr-7*.

The annotation tool is an add-on in the software package and can be used optionally to annotate the developmental stages of the individual animals. It adds the annotated information to the corresponding measurements. The annotated information includes the user defined developmental stage, but also the optional body or gonad arm length. The body and gonad arm length can be measured within the add-on itself, in which the user pre-defines the pixel to micrometer ratio corresponding to the set-up used for each image.

2.3.4 User interface and workflow of the WormFISH package

In this section, we explain in detail the graphical user interface and the workflow of the WormFISH package. First, the experiment folder has to be assigned in which the animals belonging to one batch that have to be analyzed are located (Fig. 4A). The multiple paths of one experimental batch are loaded and can be accessed at any point in the analysis process. Also, multiple channels can be loaded and accessed at any point to aid, for example, the assignment of the ROI based on the GFP channel.

After assigning the experimental folder and loading in the chosen channels, the workflow for identifying the fluorescent spots starts (Fig. 4B). The following analysis steps have to be carried out by the user: 1) assigning the POI(s) for each individual animal, 2) assigning the ROI(s) for each assigned POI, 3) Setting the threshold for each fluorophore channels with aid of spot location within the ROIs, the fluorescent spot count versus threshold graph and the fluorescent spot brightness histogram, with the option to assign the threshold within one animal differently for each ROI, 4) automatic counting of fluorescent spots of each ROI for the set threshold for all ROIs in each animal.

During the entire analysis process for spot identification, the user can start the annotation tool to add annotations to each of the measurements (Fig. 4C). The

annotation tool opens in a new window in which the user can start annotating information to each of the animals in the batch or experiment. It offers the option to open images and measure in those lengths of the body or the gonad, with the user having to provide the calibration of pixel size in micrometers for the microscopy setup used.

The quantified fluorescent spot count for each of the ROIs is saved in a simplified CSV file, including the information on the label, animal and channel of each ROI. Once the smFISH analysis has been completed, the results are saved as a *json* file and a CSV file. The pickle file contains of each ROI the number of fluorescent spots in combination with the animal name, the label of the POI, the set threshold, channel information and optional annotation data. In contrast to the CSV file, the *json* file also contains information on the average brightness of the fluorescent spots of each ROI, and individual location and brightness per spot. Since the program can be saved at any time, it allows the user to resume the analysis at any point during the analysis process.

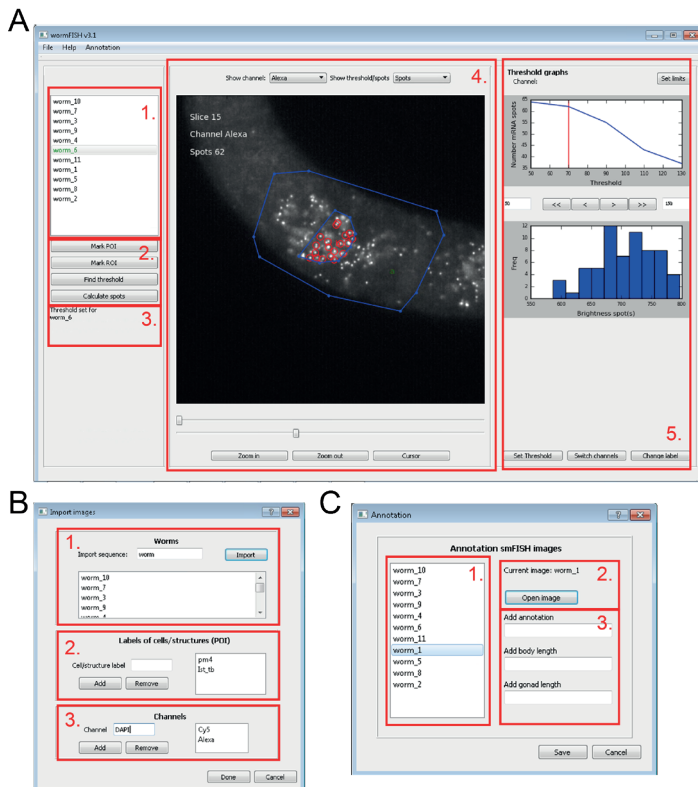


Figure 4. Screenshots workflow WormFISH. (A) Main window of the WormFISH software started when running the script for the first time. The main window shows a list of all smFISH image stacks within the loaded batch or experiment (1). Underneath it has four buttons to start each of the subsequent analysis steps (2) and underneath the last button is a panel that gives feedback on the analysis progress (3). The middle panel shows the current slice of the smFISH image stack and during each analysis step the required user input changes automatically (4). The user can change the contrast, brightness, channels, and the visualization of the spots or threshold during the threshold step. The left panel shows during the threshold step on top the threshold curve with custom range, and below the brightness histogram (5). (B) The smFISH data is loaded in a separate load window that is accessible under "File" -> "Open new" (A). The user provides information on smFISH image stacks that are analyzed, such as file names (1), POI labels (2) and fluorescence channels of interest (3). (C) The annotation tool can be started up at any point during the analysis process via the menu in the main window (A). General annotation on developmental stage, body length and gonad length (3) can be added for each smFISH image using the form (1). Images of the animal's anatomy can be loaded in a separate window to measure, for example the body or gonad length (2).

2.3.5 Quality of the fluorescent spot identification

To test the quality of smFISH spot quantification, we benchmarked WormFISH against the widely used FISH-quant package (Mueller et al. 2013). Like WormFISH, FISH-quant is not fully automated and requires user input for choosing the threshold for spot identification. In contrast to our WormFISH package, FISH-quant offers a build-in analysis of nascent mRNAs at the transcription sites, providing a more accurate estimation of the mRNA levels. To compare only the basic algorithm responsible for identifying fluorescent spots of both software packages, we excluded the analysis steps of FISH-quant for nascent mRNA quantification and focused only on counting mature mRNA molecules.

For benchmarking, we compared different datasets of L2/L3 staged animals, where the same gene (*gcy-14*, expressed in the ASEL neuron) is stained by Cy5- or Alexa594-labeled smFISH probes. The limited cell region and the relatively high expression levels of the gene resulted in frequent overlap of spots. These features in the data offer the possibility to examine how well the software packages identify overlapping spots and test their performance on images of different signal quality, since Cy5 data has a higher contrast between spot and background than Alexa594. When comparing spot identification quality between FISH-quant and WormFISH, we chose the threshold based on two equally important arguments: 1) if there is a plateau in the spot count versus thresholds curve, the threshold at the start of the plateau is chosen, 2) a threshold is chosen that results in the least amount of false positive spots (preferably none) and the highest amount of true positive spots based on the location of identified spots observed by eye. Finally, as a ground truth to compare the results of both packages against, we also counted the number of fluorescent spots in the Cy5 and Alexa594 smFISH images by hand.

The total number of samples was 20 for both the Cy5 and Alexa594 probe sets. The accuracy, sensitivity and precision are calculated with the spots identified by the software versus the spots counted by eye, in which we disguise between correctly classified spots (True Positive, TP), misclassified spots (False Positive, FP) and spots that were missed (False Negative, FN). WormFISH and FISH-quant performed similarly in identifying the fluorescent spots of Cy5 and Alexa598 smFISH data, with an average accuracy score, $TP/(TP+FP+FN)$, between 0.91-0.93 (Fig. 5A-B). Both software packages performed slightly worse on the Alexa594 smFISH data, likely because the fluorescent spot quality with Alexa598 is worse. In general, both software packages had a slight underestimation of the number of fluorescent spots. Moreover, decreasing the threshold in both packages led to a higher number of false positives (data not shown). Most false negatives were overlapping spots that both FISH-quant and WormFISH were not able to identify as two individual spots. Both software packages could identify the individual spots in closely overlapping spots between 56-65% of the cases. Overall, the WormFISH software package performed equally well in identifying spots, even when spots overlapped spatially or when the quality of the fluorescent spots was suboptimal.

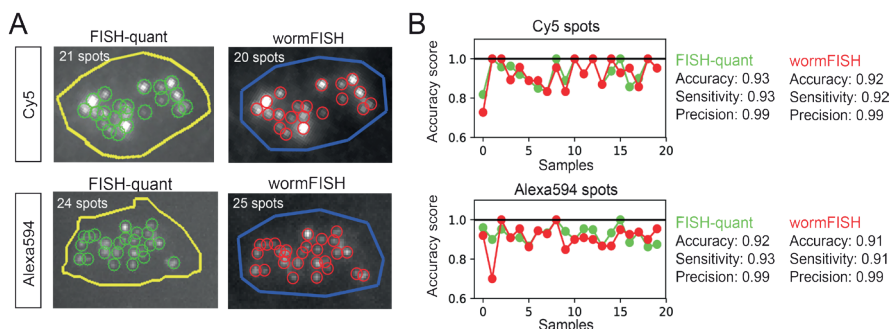


Figure 5. Spot identification quality of WormFISH similar to FISH-quant. (A) Maximum projection smFISH images of identified fluorescent spots by the software package FISH-quant and WormFISH. In the smFISH images *gcy-14* is tagged with Cy5 and Alexa594 probes. (B) Comparison of spot identification quality between software packages FISH-quant and WormFISH with Cy5 and Alexa594 on *gcy-14* in L2/L3 staged animals. The accuracy score is calculated per sample ($n=20$). The average accuracy, sensitivity and precision scores are shown in the right panel for both FISH-quant and WormFISH.

2.4 Discussion

We have developed a software packages, WormFISH, for counting smFISH spots, specifically aimed at large *C. elegans* smFISH data sets and combined with tools to annotate developmental stage information. In the WormFISH package, the workflow is optimized to deal with challenges specific to *C. elegans* smFISH data

sets. First, such data sets consist out a multiple smFISH image stacks containing a single animal, in which typically only a few regions or cells of interest are located. Most smFISH analysis software packages lack an analysis step in which the user is asked to specify the points of interest (POIs) prior to specifying the regions of interest (ROIs). The added POI step in WormFISH, in which the software zooms in on each POI during setting the ROI, shortens the time spend on annotating ROIs in smFISH images. Another challenge is the varying smFISH spot quality between animals, making batch automation impossible. To aid in enhanced batch processing, the WormFISH package allows the user to easily switch between individual animals in one batch during the entire analysis process, to allow for fast comparison between animals. The last significant challenge that comes with analyzing *C. elegans* smFISH data, is the varying quality of smFISH spots between image slices, due light scattering within the animal's tissue, particularly in older animals with a comparatively large size. The WormFISH software package offers the option to assign the spot detection threshold within one animal differently for each ROI.

The spot identification algorithm in WormFISH is based on the previously described algorithm of Raj et al. (Raj et al. 2008), with an added 3D local peak algorithm after the convolution and filtering to aid in separating overlapping spots. The WormFISH package has a relatively basic algorithm for identifying the fluorescent spots in the WormFISH software package, compared to more extensive smFISH analysis software packages, such as the widely-used FISH-QUANT software package (Mueller et al. 2013). Nevertheless, the WormFISH software package preformed equally well in identifying smFISH spots compared to FISH-quant software package, even when spots overlap or when the quality of the fluorescent spots is less optimal.

The rate-limiting step in time spend on analyzing *C. elegans* smFISH data, is setting the threshold by hand for each individual animal and optionally each ROI. In addition, the spot identification algorithm is relatively slow, due to the incorporation of the local peak algorithm. There are software packages that offer automated thresholding to increase the workflow, for example utilizing machine learning for identifying smFISH spots (Rifkin 2011). However, those are limited in their performance when analyzing *C. elegans* smFISH data with varying quality of smFISH spots, both within a single animal and between animals in one batch. Nevertheless, it would be interesting to explore the possibility of using machine learning algorithms aiding in the identification of smFISH spots, to resolve the rate-limiting step of manual thresholding.

Studying gene expression dynamics in time with smFISH experiments is a challenge, since the samples need to be fixed. However, *C. elegans* development is

highly invariant (Sulston and Horvitz 1977), offering the possibility to stage animals according to their morphology and anatomical features, such as staging of L4 staged animals by their vulva morphology (Mok, Sternberg, and Inoue 2015). We added an annotation tool to the software package, that enables the user to assigning precise developmental stages to each smFISH measurement. The addition of precise developmental stage proved to be useful in the case of oscillating gene expression of *nhr-7*, in which the vulva/gonad/sperm morphology offered more insight into the dynamics of the expression levels during the development (Fig. 3C). Taken together, the WormFISH package offers a specific solution for analyzing large *C. elegans* smFISH data sets on dynamics in expression during development, while delivering a reliable quantification of the smFISH fluorescent spots.

2.5 Material and methods

Single molecule fluorescence in situ hybridization

The oligonucleotides for the smFISH probe sets were designed with optimal GC content and specificity for the gene of interest using the Stellaris RNA FISH probe designer. The oligonucleotides were synthesized with a 3' amino C7 modification and purified by LGC Biosearch Technologies. Conjugation of the oligonucleotides with either Cy5 (GE Amersham) or Alexa594 (Invitrogen) was done as previously described [24]. The smFISH protocol was performed as previously described [2, 16]. Briefly, staged animals were washed from plates with M9 buffer and fixed in 4% formaldehyde in 1x PBS, gently rocking at room temperature (RT) for 40 minutes (young adults for 35 minutes). Fixation of embryos required a snap-freeze step to crack the eggshells by submerging embryos, after 15 minutes in fixation buffer, in liquid nitrogen, and thawing on ice for 20 minutes. After fixation, the animals were 2x washed with 1xPBS and resuspended in 70% ethanol overnight at 4°C. Ethanol was removed and animals were washed with 10% formamide and 2X SSC, as preparation for the hybridization. Animals were incubated with the smFISH probes overnight in the dark at 37°C in a hybridization solution (Stellaris) with added 10% formamide. The next day, animals were washed 2x with 10% formamide and 2X SSC each with an incubation step of 30 minutes at 37°C. The last wash step contains DAPI 5 µg/mL for nuclear staining. The wash buffer was removed, and animals were resuspended in 2X SSC and stored at 4°C until imaging. The 2X SSC was aspirated and animals were immersed in 100 µl GLOX buffer (0.4% glucose, 10 mM Tris-HCl, pH 8.0, 2X SSC) together with 1 µl Catalase (Sigma-Aldrich) and 1 µl glucose oxidase (Sigma-Aldrich) (3.7 mg/mL) to prevent bleaching during imaging.

Microscopy images of smFISH samples were acquired with a Nikon Ti-E inverted fluorescence microscope, equipped with a 100X plan-apochromat oil-immersion

objective and an Andor Ikon-M CCD camera controlled by μ Manager software [25]. The smFISH analysis was performed with the WormFISH software package

2.6 Appendix

2.6.1 Author contributions to the chapter

JJHT and JSvZ conceptualized the design. JJHT wrote the python scripts. JJHT performed the experiments and analyzed the data. JJHT wrote the text. JSvZ reviewed and edited the text. JSvZ supervised the study.

2.6.2 Supplementary figures

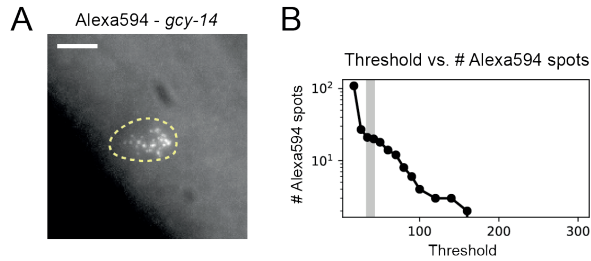


Figure S1: Threshold versus number of identified Alexa594 fluorescent spots. (A) Maximum projection of a Alexa594 smFISH image stack (*gcy-14* mRNA, ASEL neuron), with the corresponding region of interest (ROI). (B) Threshold versus the number of counted Alexa594 fluorescent spots, quantified from the example image in (A). Alexa594 shows a less well-defined plateau. Grey inset highlights the chosen threshold by the user.

2.7 References

- Byerly, L., R. C. Cassada, and R. L. Russell. 1976. 'The life cycle of the nematode *Caenorhabditis elegans*. I. Wild-type growth and reproduction,' *Dev Biol*, 51: 23-33.
- Chong, S., C. Chen, H. Ge, and X. S. Xie. 2014. 'Mechanism of transcriptional bursting in bacteria,' *Cell*, 158: 314-26.
- Femino, A. M., F. S. Fay, K. Fogarty, and R. H. Singer. 1998. 'Visualization of single RNA transcripts in situ,' *Science*, 280: 585-90.
- Frand, A. R., S. Russel, and G. Ruvkun. 2005. 'Functional genomic analysis of *C. elegans* molting,' *PLoS Biol*, 3: e312.
- Frokjaer-Jensen, C., N. Jain, L. Hansen, M. W. Davis, Y. Li, D. Zhao, K. Rebora, J. R. M. Millet, X. Liu, S. K. Kim, D. Dupuy, E. M. Jorgensen, and A. Z. Fire. 2016. 'An Abundant Class of Non-coding DNA Can Prevent Stochastic Gene Silencing in the *C. elegans* Germline,' *Cell*, 166: 343-57.
- Gritti, N., S. Kienle, O. Filina, and J. S. van Zon. 2016. 'Long-term time-lapse microscopy of *C. elegans* post-embryonic development,' *Nat Commun*, 7: 12500.
- Halpern, K. B., R. Shenhav, O. Matcovitch-Natan, B. Toth, D. Lemze, M. Golan, E. E. Massasa, S. Baydatch, S. Landen, A. E. Moor, A. Brandis, A. Giladi, A. S. Avihail, E. David, I. Amit, and S. Itzkovitz. 2017. 'Single-cell spatial reconstruction reveals global division of labour in the mammalian liver,' *Nature*, 542: 352-56.
- Hansen, M. M. K., R. V. Desai, M. L. Simpson, and L. S. Weinberger. 2018. 'Cytoplasmic Amplification of Transcriptional Noise Generates Substantial Cell-to-Cell Variability,' *Cell Syst*, 7: 384-97 e6.
- Harterink, M., D. H. Kim, T. C. Middelkoop, T. D. Doan, A. van Oudenaarden, and H. C. Korswagen. 2011. 'Neuroblast migration along the anteroposterior axis of *C. elegans* is controlled by opposing gradients of Wnts and a secreted Frizzled-related protein,' *Development*, 138: 2915-24.
- Huelsz-Prince, G., and J. S. van Zon. 2017. 'Canalization of *C. elegans* Vulva Induction against Anatomical Variability,' *Cell Syst*, 4: 219-30 e6.
- Ji, N., and A. van Oudenaarden. 2012. 'Single molecule fluorescent in situ hybridization (smFISH) of *C. elegans* worms and embryos,' *WormBook*: 1-16.
- Killian, D. J., and E. J. Hubbard. 2005. 'Caenorhabditis elegans germline patterning requires coordinated development of the somatic gonadal sheath and the germ line,' *Dev Biol*, 279: 322-35.
- Kimble, J. E., and J. G. White. 1981. 'On the control of germ cell development in *Caenorhabditis elegans*,' *Dev Biol*, 81: 208-19.
- Kimble, J., and D. Hirsh. 1979. 'The postembryonic cell lineages of the hermaphrodite and male gonads in *Caenorhabditis elegans*,' *Dev Biol*, 70: 396-417.
- Lee, C., E. B. Sorensen, T. R. Lynch, and J. Kimble. 2016. '*C. elegans* GLP-1/Notch activates transcription in a probability gradient across the germline stem cell pool,' *Elife*, 5.
- Little, S. C., M. Tikhonov, and T. Gregor. 2013. 'Precise developmental gene expression arises from globally stochastic transcriptional activity,' *Cell*, 154: 789-800.
- Lyubimova, A., S. Itzkovitz, J. P. Junker, Z. P. Fan, X. Wu, and A. van Oudenaarden. 2013. 'Single-molecule mRNA detection and counting in mammalian tissue,' *Nat Protoc*, 8: 1743-58.
- Meeuse, M. W., Y. P. Hauser, L. J. Morales Moya, G. J. Hendriks, J. Eglinger, G. Bogaarts, C. Tsiairis, and H. Grosshans. 2020. 'Developmental function and state transitions of a gene expression oscillator in *Caenorhabditis elegans*,' *Mol Syst Biol*, 16: e9498.
- Middelkoop, T. C., L. Williams, P. T. Yang, J. Luchtenberg, M. C. Betist, N. Ji, A. van Oudenaarden, C. Kenyon, and H. C. Korswagen. 2012. 'The thrombospondin repeat

- containing protein MIG-21 controls a left-right asymmetric Wnt signaling response in migrating *C. elegans* neuroblasts', *Dev Biol*, 361: 338-48.
- Mok, D. Z., P. W. Sternberg, and T. Inoue. 2015. 'Morphologically defined sub-stages of *C. elegans* vulval development in the fourth larval stage', *BMC Dev Biol*, 15: 26.
- Mueller, F., A. Senecal, K. Tantale, H. Marie-Nelly, N. Ly, O. Collin, E. Basyuk, E. Bertrand, X. Darzacq, and C. Zimmer. 2013. 'FISH-quant: automatic counting of transcripts in 3D FISH images', *Nat Methods*, 10: 277-8.
- Omerzu, M., N. Fenderico, B. de Barbanson, J. Sprangers, J. de Ridder, and M. M. Maurice. 2019. 'Three-dimensional analysis of single molecule FISH in human colon organoids', *Biol Open*, 8.
- Pharris, M. C., T. C. Wu, X. Chen, X. Wang, D. M. Umulis, V. M. Weake, and T. L. Kinzer-Ursem. 2017. 'An automated workflow for quantifying RNA transcripts in individual cells in large data-sets', *MethodsX*, 4: 279-88.
- Raj, A., C. S. Peskin, D. Tranchina, D. Y. Vargas, and S. Tyagi. 2006. 'Stochastic mRNA synthesis in mammalian cells', *PLoS Biol*, 4: e309.
- Raj, A., P. van den Bogaard, S. A. Rifkin, A. van Oudenaarden, and S. Tyagi. 2008. 'Imaging individual mRNA molecules using multiple singly labeled probes', *Nat Methods*, 5: 877-9.
- Raj, A., and A. van Oudenaarden. 2008. 'Nature, nurture, or chance: stochastic gene expression and its consequences', *Cell*, 135: 216-26.
- Rifkin, S. A. 2011. 'Identifying fluorescently labeled single molecules in image stacks using machine learning', *Methods Mol Biol*, 772: 329-48.
- Sanchez, A., and I. Golding. 2013. 'Genetic determinants and cellular constraints in noisy gene expression', *Science*, 342: 1188-93.
- Stoeger, T., N. Battich, M. D. Herrmann, Y. Yakimovich, and L. Pelkmans. 2015. 'Computer vision for image-based transcriptomics', *Methods*, 85: 44-53.
- Sulston, J. E., and H. R. Horvitz. 1977. 'Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans*', *Dev Biol*, 56: 110-56.
- Topalidou, I., A. van Oudenaarden, and M. Chalfie. 2011. 'Caenorhabditis elegans aristaless/Arx gene *alr-1* restricts variable gene expression', *Proc Natl Acad Sci U S A*, 108: 4063-8.
- Tsanov, N., A. Samacoits, R. Chouaib, A. M. Traboulsi, T. Gostan, C. Weber, C. Zimmer, K. Zibara, T. Walter, M. Peter, E. Bertrand, and F. Mueller. 2016. 'smiFISH and FISH-quant - a flexible single RNA detection approach with super-resolution capability', *Nucleic Acids Res*, 44: e165.
- Uppaluri, S., and C. P. Brangwynne. 2015. 'A size threshold governs *Caenorhabditis elegans* developmental progression', *Proc Biol Sci*, 282: 20151283.
- van der Walt, S., J. L. Schonberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, T. Yu, and contributors scikit-image. 2014. 'scikit-image: image processing in Python', *PeerJ*, 2: e453.

Chapter 3

3

Mechanism of life-long maintenance of neuron identity despite molecular fluctuations

Joleen J. H. Traets¹, Servaas N. van der Burght^{2,3}, Gert Jansen², Jeroen S. van Zon¹

¹AMOLF, Science Park 104, 1098 XG Amsterdam, the Netherlands

²Department of Cell Biology, Erasmus University Medical Centre,
Dr. Molewaterplein 40, 3015 GD, Rotterdam, the Netherlands

³Present address: The Gurdon Institute, University of Cambridge,
Cambridge CB2 1QN, UK

(Submitted for publication)

3.1 Abstract

Cell fate is maintained over long timescales, yet molecular fluctuations can lead to spontaneous loss of this differentiated state. We uncovered a mechanism that explains life-long maintenance of ASE neuron fate in *C. elegans* by the terminal selector transcription factor CHE-1. Fluctuations in CHE-1 level are buffered by the reservoir of CHE-1 bound at its target promoters, which ensure continued *che-1* expression by preferentially binding the *che-1* promoter. We validated this mechanism by showing that *che-1* expression was resilient to induced transient CHE-1 depletion, while both expression of CHE-1 targets and ASE function were lost. We identified a 130 bp *che-1* promoter fragment responsible for this resilience, with deletion of a homeodomain binding site in this fragment causing stochastic loss of ASE identity long after its determination. Because network architectures that support this mechanism are highly conserved in cell differentiation, it may explain stable cell fate maintenance in many systems.

3.2 Introduction

In most animal tissues, terminally differentiated cells are renewed on timescales of days to months (Leblond and Walker 1956), yet the nervous system is unique, as most neurons hardly renew at all (Ming and Song 2005). Many mature neuron subtypes exist that differ in the expression of hundreds of subtype-specific genes (Deneris and Hobert 2014). How neuronal cells maintain this terminally differentiated state over such long timescales - decades, in the case of humans - is not understood.

Recent studies have found that differentiation of neuron subtypes is often controlled by only 1 or 2 different transcription factors, called “terminal selectors” (Hobert 2016). These act through a conserved network motif called a single-input module (Alon 2007): they bind to specific *cis*-regulatory control elements to induce both their own expression and that of the downstream target genes that define the neuronal subtype (Fig. 1A). Such terminal selector networks have been found to underlie differentiation of many neuron types in the nematode *Caenorhabditis elegans* (Deneris and Hobert 2014; Hobert 2016), photoreceptor subtypes in *Drosophila* (Hsiao et al. 2013) and dopaminergic neurons in mice (Ninkovic et al. 2010), indicating that they form an evolutionary conserved principle for neuron subtype determination.

Terminal selectors positively regulate their own expression, raising the possibility that they act as bistable genetic switches. Such switches are wide-spread in biology (Ferrell 2002; Alon 2007) and are often seen as an attractive mechanism to explain cell fate determination. In this hypothesis, at the time of determination, transient signals induce expression of terminal selectors, which then maintain their own expression, and that of all target genes, by autoregulation in the subsequent absence of these signals (Fig. 1B) (Hobert 2008). However, a key weakness of bistable switches is that they remain reversible at all times, with a transient decrease in terminal selector levels potentially sufficient to lose terminal selector expression and, presumably, cell fate (Fig. 1B). Indeed, bistable genetic switches often suffer from stochastic transitions between their different states due to molecular noise, i.e. random fluctuations in the levels of their core components (Ozbudak et al. 2004; Suel et al. 2006; Acar, Becskei, and van Oudenaarden 2005). Hence, it is often assumed that stable cell fate maintenance must require additional feedback mechanisms, such as histone or chromatin modifications (Orlando 2003; Ringrose and Paro 2007), that lock-in cell fate in an essentially irreversible manner.

Here, we studied long-term maintenance of neuron subtype in the salt-sensing ASE neurons of the nematode *C. elegans* (Fig. 1A). ASE subtype is controlled by the terminal selector CHE-1, a transcription factor whose expression is transiently

induced by the nuclear hormone receptor NHR-67 at the time of determination (Sarin et al. 2009). CHE-1 induces the expression of 500-1000 ASE-specific target genes, such as chemosensory receptors, ion-channels, and neuropeptides, by binding ASE motifs within their promoters (Etchberger et al. 2007). Its continued presence is required for expression of target genes after subtype determination (Etchberger et al. 2009). CHE-1 also upregulates its own expression. This positive feedback loop is necessary for sustaining *che-1* expression and ASE cell fate directly after cell determination (Etchberger et al. 2007; Leyva-Diaz and Hobert 2019). However, whether this positive feedback loop by itself is sufficient to ensure life-long maintenance of ASE fate is unknown. The impact of molecular noise, such as variability in CHE-1 protein copy number, on ASE fate maintenance has not been studied. Overall, it is an open question whether a reversible, bistable switch based on positive CHE-1 autoregulation would be sufficiently stable to maintain ASE fate for the animal's lifetime, or if additional mechanisms are necessary to ensure that, once ASE fate is determined, *che-1* expression becomes independent of CHE-1 itself and can no longer spontaneously switch off.

We show that sufficiently long, transiently induced depletion of CHE-1 caused permanent loss of ASE fate, indicating that it is controlled by a switch that remains reversible long after specification. This raises the question how the switch is protected against molecular noise, which could cause it to spontaneously lose ASE fate. Combining experimental measurements of the key parameters that control the magnitude of noise, i.e. the copy numbers and lifetimes of *che-1* mRNA and protein, with stochastic models of the *che-1* genetic network, revealed a novel mechanism, “target reservoir buffering”, that dramatically increased switch stability. This stability resulted from the presence of a reservoir of CHE-1 protein bound at the promoters of its target genes, coupled with preferential binding of CHE-1 to the *che-1* promoter compared to the promoters of its other targets. This led to exceedingly stable ON states (high *che-1* expression), with spontaneous transitions to the OFF state (no *che-1* expression) observed at rates of <1 /year. Consistent with this mechanism, we found that upon induced CHE-1 depletion *in vivo*, *che-1* mRNA expression remained present, even when expression of other target genes vanished together with the animal's ability to respond to salt. This allowed full recovery of CHE-1 protein levels, target gene expression and chemosensation if induced CHE-1 depletion was sufficiently short. We found a 130 bp promoter region surrounding the *che-1* ASE motif responsible for this resilience of *che-1* expression to CHE-1 depletion. This region contained a homeodomain protein binding site that, when mutated, caused stochastic loss of ASE fate well after ASE specification, indicating a strong decrease in stability of the ON state. We therefore speculate that homeodomain proteins have a role in maintaining and stabilizing ASE fate, potentially by recruiting CHE-1 preferentially to its own promoter.

3.3 Results

3.3.1 Loss of ASE neuron fate upon transient CHE-1 depletion

To test whether positive autoregulation of *che-1* expression is necessary for ASE fate maintenance, we depleted CHE-1 protein levels in ASE neurons *in vivo*, using the auxin inducible degradation system (Zhang et al. 2015; Serrano-Saiz et al. 2018). *che-1::GFP::AID* animals were exposed to 1 mM auxin to induce CHE-1 depletion (Fig. 1C). CHE-1::GFP was strongly reduced after ~1.5 hr, and undetectable after ~3 hrs exposure to auxin (Fig. 1D, Fig. S1A). To examine how CHE-1 depletion impacted ASE function, we used a quadrant chemotaxis assay to quantify the chemotaxis response of CHE-1 depleted animals (Wicks et al. 2000; Jansen, Weinkove, and Plasterk 2002). In this assay, animals choose between two agar quadrants with NaCl and two without (Fig. S1B). *che-1::GFP::AID* animals exposed to auxin for 24 hrs showed reduced chemotaxis to NaCl ($P < 0.001$) (Fig. 1E, S1C). This agrees with previous results showing that permanent inhibition of *che-1* expression by RNAi in larvae resulted in reduced expression of CHE-1 target genes (Etchberger et al. 2009). However, by inhibiting *che-1* expression permanently, rather than transiently, these results left open the possibility that, after initial ASE subtype determination, *che-1* expression is maintained independently of CHE-1 in an irreversible manner.

In contrast, if *che-1* expression is bistable and reversible, then a sufficiently long period of transient CHE-1 depletion should result in permanent loss of *che-1* expression (Fig. 1B). To test this, we exposed animals to auxin for increasing time intervals (Fig. 1C), and analysed CHE-1::GFP expression and NaCl chemotaxis after 24 (Fig. 1D, S1B,C) or 48 hrs (Fig. 1E,F) recovery on plates without auxin. In animals exposed to auxin for 24 hrs, both CHE-1::GFP expression and NaCl chemotaxis returned to wild-type levels after 24 hrs recovery. However, after 48 hrs auxin exposure, CHE-1::GFP did not return in 11/41 animals (Fig. 1F, S1B). This fraction further decreased with the duration of auxin exposure, with 0/9 animals recovering CHE-1::GFP expression after 120 hrs on auxin. Similarly, the chemotaxis response did not recover in animals exposed to auxin for 96 hrs or longer (Fig. 1E, S1C; Table S1), suggesting that the absence of CHE-1::GFP resulted in loss of ASE subtype identity. Overall, these observations indicated that CHE-1 controls its expression, and ASE fate, as a bistable switch that, even though it can withstand strong decreases in CHE-1 level for a limited time, remains fully reversible long after its induction.

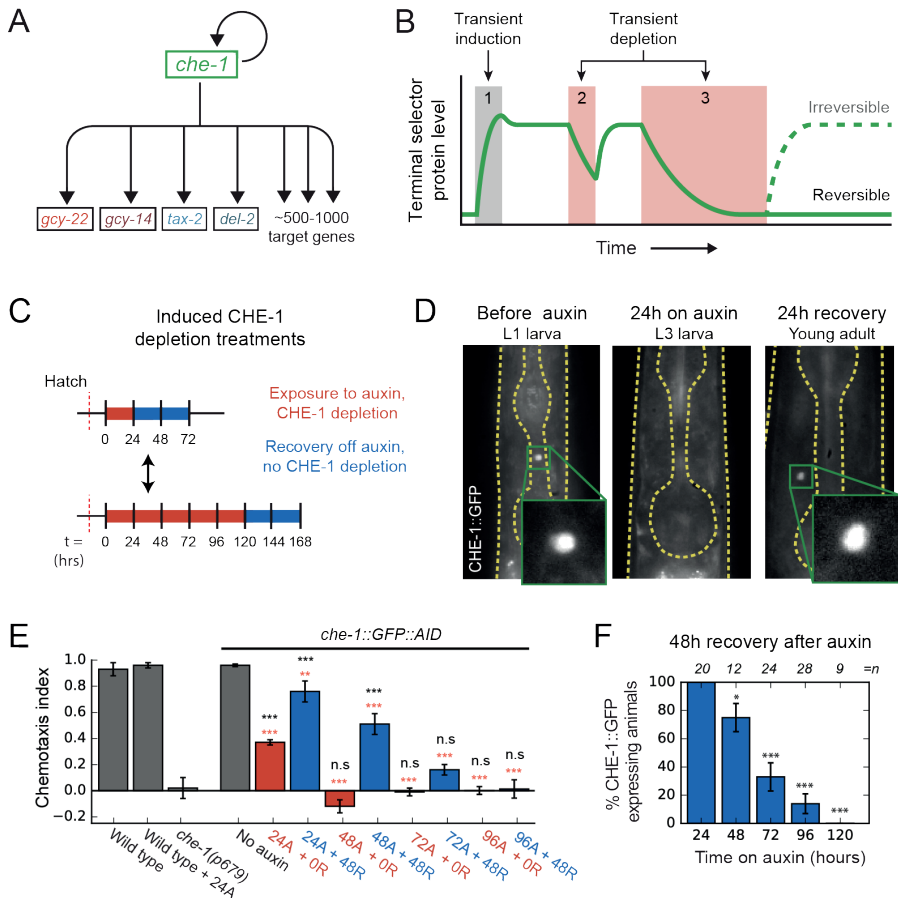


Figure 1. Loss of ASE neuron fate upon transient CHE-1 depletion. (A) The terminal selector gene *che-1* induces its own expression and that of 500-1000 target genes that together determine ASE neuron fate. Positive autoregulation of *che-1* expression could result in bistable, switch-like behaviour. (B) Bistability generates sustained terminal selector expression upon transient induction (1), that is resilient to short periods of terminal selector depletion (2). However, bistable switches remain reversible and will lose terminal selector expression upon sufficiently long depletion (3), while irreversible switches will always recover. (C) Transient CHE-1 depletion using Auxin-Induced Degradation (AID). *che-1::GFP::AID* L1 larvae (CHE-1::GFP) or young adults (chemotaxis) were exposed for different time periods to 1 mM auxin to induce CHE-1 degradation, and were subsequently characterized after a 24 or 48 hrs recovery period. (D) CHE-1::GFP fluorescence in *che-1::GFP::AID* animals before (left) and after 24 hrs auxin treatment (middle), and after a subsequent 24 hrs recovery off auxin (right). Even though CHE-1::GFP is lost from ASE neurons after auxin treatment, it reappears after recovery off auxin. (E) Response to 10 mM NaCl for wild-type animals, *che-1(p679)* mutants defective in NaCl chemotaxis, and *che-1::GFP::AID* animals exposed to auxin for 24-96 hrs (24A – 96A) tested directly (0R) or after 48 hrs recovery (48R). *che-1::GFP::AID* animals on auxin showed a chemotaxis defect similar to *che-1(p679)* mutants. *che-1::GFP::AID* animals recovered chemotaxis to NaCl after 24 or 48 hrs on auxin, but exhibited a persistent chemotaxis defect after sufficiently long, transient CHE-1::GFP::AID depletion. (F) Fraction of animals that

recovered CHE-1::GFP expression 48 hrs after auxin treatment of increasing length. No animals recovered CHE-1::GFP expression after 120 hrs of CHE-1 depletion. Error bars in E and F represent mean of 4 assays \pm S.E.M. *P<0.05, **P<0.01, ***P<0.001, in E significance is compared to *che-1(p679)* mutants (black) or no auxin (red).

3.3.2 Copy number and lifetime of *che-1* mRNA and protein

The observed reversibility of the ON state raises the question how spontaneous transitions to the OFF state, due to random fluctuations in CHE-1 level, are prevented under normal conditions. Theoretical studies of genetic switches suggest that the probability of such transitions decreases with increasing average copy number and lifetime of the transcription factors involved (Warren and ten Wolde 2005; Walczak, Onuchic, and Wolynes 2005; Mehta, Mukhopadhyay, and Wingreen 2008). Therefore, we determined copy numbers and lifetimes of *che-1* mRNA and protein.

First, we measured the absolute number of *che-1* mRNA and protein molecules in ASE neurons. We used single-molecule FISH (smFISH) (Ji and van Oudenaarden 2012) to count individual *che-1* mRNA molecules. As a benchmark, we also measured mRNA levels of the putative NaCl receptors *gcy-14* and *gcy-22*, and the ion channel subunits *del-2* and *tax-2*, which are CHE-1 target genes with >1 ASE motif in their promoter region (Ortiz et al. 2009; Coburn and Bargmann 1996). In embryos, we found that *che-1* mRNA levels peaked at 26 ± 6 mRNAs/cell during the bean stage, i.e. the time of ASE neuron determination (Fig. 2A,B), and fell to 6 ± 3 mRNAs/cell from the comma stage onwards. In contrast, the CHE-1 target gene *gcy-22* showed a steady increase in mRNA levels during development and surpassed *che-1* mRNA expression after the 1.5-fold stage. In larvae, we also found low *che-1* mRNA levels, with 5 ± 2 mRNAs in the left (ASEL) and 7 ± 3 mRNAs in the right ASE neuron (ASER). *che-1* expression was low compared to the panel of CHE-1 target genes examined (Fig. 2C,D), with *del-2*, *gcy-14* and *gcy-22* expressed at significantly higher levels. Moreover, *che-1* mRNA levels remained low during all four larval stages, L1-L4, with slightly higher mRNA levels in the ASER, compared to the ASEL neuron (Fig. 2E). For this entire period, the target genes with the highest expression, *gcy-14* and *gcy-22*, remained expressed well above *che-1* levels. Overall, these results show that *che-1* expression is biphasic: an initiation phase with high *che-1* expression level when ASE fate is induced in the embryo, followed by a maintenance phase with *che-1* expressed at a low level where molecular copy number fluctuations likely have significant impact.

We also estimated the absolute CHE-1 protein copy number in the ASE neurons. We immersed endogenously tagged *che-1::GFP* animals (Leyva-Diaz and

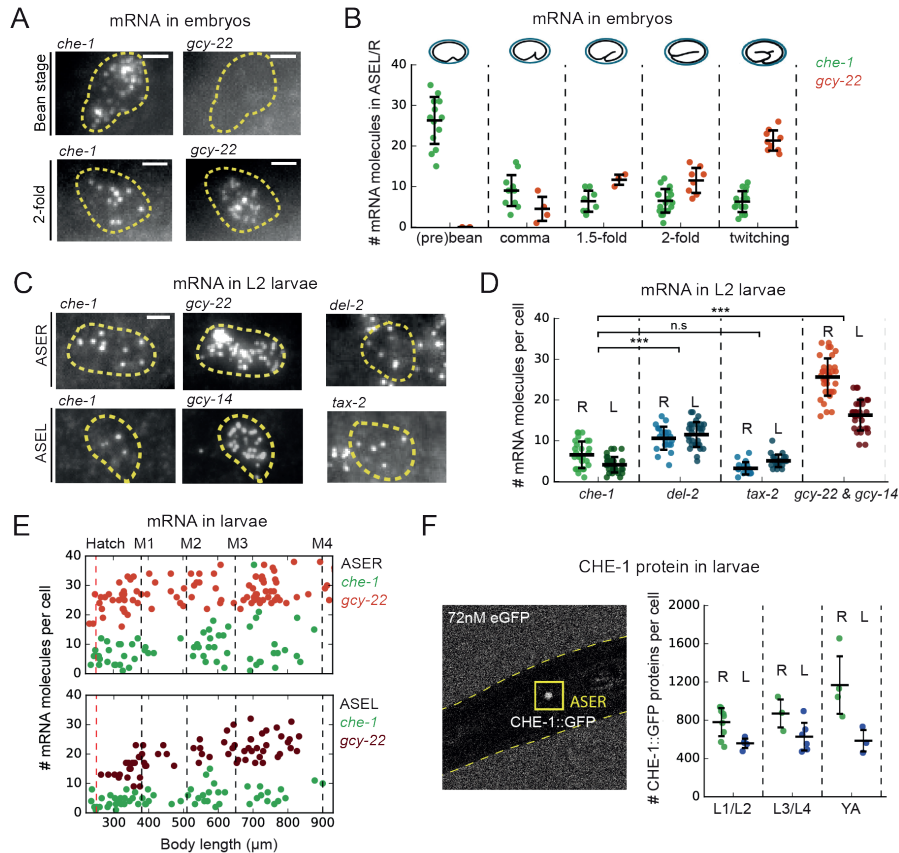


Figure 2. *che-1* mRNA and protein copy numbers in ASE neurons. (A) Expression of *che-1* and the CHE-1 target *gcy-22* in embryos at the bean (top) and 2-fold stage (bottom). Each spot is a single mRNA molecule visualized by single molecule FISH (smFISH). Dashed lines outline ASE neuron cell bodies. Scale bar: 2 mm. (B) *che-1* (green) and *gcy-22* (red) mRNA levels during embryonic development. *che-1* expression peaks during specification, but falls as expression of CHE-1 target genes rise. (C) Expression of *che-1* and CHE-1 targets *gcy-22*, *gcy-14*, *tax-2* and *del-2* visualized by smFISH in L2 larvae. Dashed lines outline left and right ASE neurons (ASEL/R). Scalebar: 2 mm. (D) Quantification of expression of *che-1* and CHE-1 targets in ASEL (L) and ASER (R) in L2 larvae. *che-1* mRNA levels are low compared to other CHE-1 target genes. (E) Low *che-1* expression (green, ASER/L) compared to expression of CHE-1 targets *gcy-22* (red, ASER) and *gcy-14* (red, ASEL) throughout larval development. Body length corresponds to developmental time, with approximate timing of hatching and molts between larval stages L1-L4 indicated by vertical lines. (F) Left panel: two-photon microscopy image of L2 larva expressing endogenously-tagged CHE-1::GFP, immersed in 72 nM eGFP for calibration. Right panel: CHE-1::GFP protein molecules in ASER (green, R) and ASEL (blue, L) at different stages of post-embryonic development. The number of CHE-1 proteins is comparable to the predicted number of CHE-1 binding sites. Error bars in B, D and F represent mean \pm SD. ***P<0.001.

Hobert 2019) in eGFP, and calculated the CHE-1 protein number by comparing the CHE-1::GFP fluorescence inside the ASE neurons with the ambient fluorescence of eGFP (Fig. 2F) (Gregor et al. 2007). For all stages of post-embryonic development, we found an average CHE-1::GFP copy number of 900 ± 250 CHE-1 (~ 325 nM) and 600 ± 120 CHE-1 (~ 260 nM) molecules/cell for the ASER and ASEL neuron, respectively (Fig. 2F, S2A), with slightly lower copy numbers observed in the embryo (Fig. S2B). Overall, the observed range of CHE-1::GFP protein levels, 500-1400 molecules/cell, is comparable to the number of CHE-1 binding sites, 500-1000, in the promoters of CHE-1 target genes, as estimated by the number of ASE motifs detected previously (Etchberger et al. 2007) and by our analysis (see Methods), indicating that CHE-1 targets compete for a limited pool of CHE-1 protein.

Next, we measured *che-1* mRNA and protein lifetimes. To determine *che-1* mRNA lifetimes in ASE neurons, we transiently overexpressed *che-1* mRNA, using a *hsp16.41p::che-1* heat shock inducible construct (Patel and Hobert 2017). By exposing animals to 37°C for ~ 10 min, we raised *che-1* mRNA levels in the ASE neurons 5-fold, to 24 ± 4 mRNAs/cell ~ 10 min after heat shock (Fig. 3A,B). Next, we shifted animals to 20°C to return *che-1* expression to its pre-induction level and fixed animals at ~ 17 min intervals to quantify *che-1* mRNA levels using smFISH. We found that *che-1* mRNA levels returned to pre-induction values after ~ 60 min. By fitting the measured *che-1* mRNA levels to an exponential function, we obtained a *che-1* mRNA half-life of 17 ± 4 min. To measure CHE-1 protein lifetime, we used Fluorescence Recovery after Photobleaching (FRAP) in CHE-1::GFP animals. CHE-1::GFP was bleached to approximately $\sim 20\%$ of the original fluorescence level and we measured the recovery of the CHE-1::GFP signal over time, until it reached pre-bleaching levels, which occurred within ~ 3 hrs (Fig. 3C,D). To estimate CHE-1::GFP protein lifetime, we fitted an exponential recovery curve to the experimental data for each individual animal, resulting in an average measured half-life of 83 ± 20 min.

The *che-1* mRNA and protein half-lives are short compared to reported mRNA half-lives in eukaryotes (Fornasiero et al. 2018; Sharova et al. 2009) and average protein half-lives reported in young *C. elegans* adults (Dhondt et al. 2017; Schwanhaussner et al. 2011). Overall, these measured lifetimes show that *che-1* mRNA and protein turnover is rapid compared to the 2-3 week lifespan of *C. elegans* during which ASE identity should be maintained.

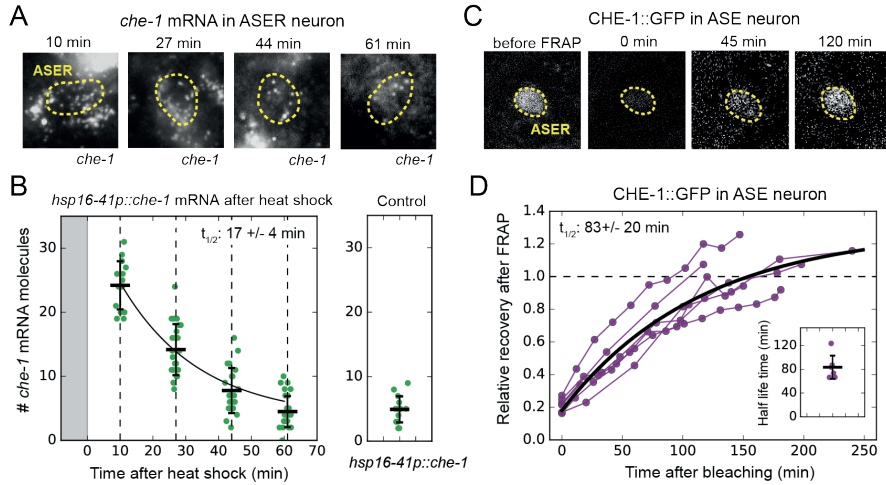


Figure 3. *che-1* mRNA and protein lifetimes. (A) *che-1* mRNAs in L2 larvae at different times after induction of *che-1* by a 37°C heat shock in *hsp16-41p::che-1* animals, visualized by smFISH. Dashed lines outline ASER neuron. (B) *che-1* mRNA level in ASE neurons of individual L2 animals (green) as function of time after a 10 min heat shock (grey area). Black line is the fitted decay curve. Control L2 larvae did not receive a heat shock. The measured *che-1* mRNA half-life was 17 ± 4 mins. (C) CHE-1::GFP fluorescence recovery after photobleaching (FRAP) in the ASER neuron of a single L4 animal. Time is indicated relative to bleaching of CHE-1::GFP. (D) Fluorescence recovery of CHE-1::GFP in ASE neurons of L4 or young adult animals (n=6). An exponential recovery curve model was fitted to data of each individual animal (black line indicates the average recovery curve). The inset shows the fitted half-life for each individual animal. The average measured CHE-1::GFP protein half-life was 80 ± 20 min. Error bars in B represent mean \pm SD. $n \geq 10$ in B and $n = 6$ in D.

3.3.3 Stochastic simulations identify stable cell fate maintenance parameters

The measurements of *che-1* mRNA and protein copy numbers and lifetimes allowed us to perform realistic simulations of the CHE-1 switch to estimate its stability against stochastic fluctuations. We constructed stochastic models that included production and decay of *che-1* mRNA and protein molecules, binding of CHE-1 to the promoter of *che-1* and target genes, and target gene expression (Fig. 4A). We examined two bistable models that differed in binding of CHE-1 to its own promoter. In the first model, we assume that CHE-1 binds as a monomer to induce *che-1* expression. This model agrees with the observation that the *che-1* promoter contains only a single ASE motif (Etchberger et al. 2007), but lacks cooperativity in *che-1* induction. Because cooperativity is considered important for generating bistability (Ferrell 2002), we also included a second model where *che-1* induction is cooperative, by assuming that expression occurs only when two CHE-1 molecules bind the *che-1* promoter.

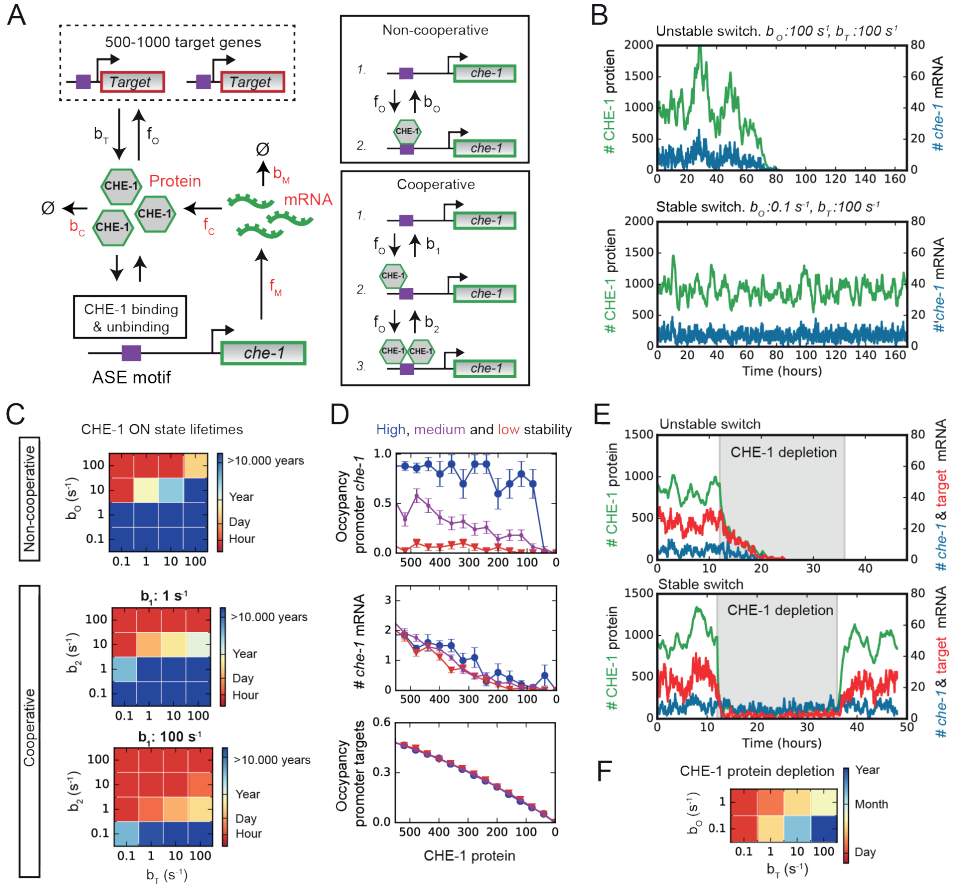


Figure 4. Stable ON state by preferential binding of CHE-1 to its own promoter. (A) Overview of the bistable, stochastic CHE-1 switch model, including production and degradation of *che-1* mRNA and protein, and binding of CHE-1 protein to its own promoter and other target genes. Parameters constrained by experiments are red. Inset: CHE-1 binding is modelled as monomers (non-cooperative) or dimers (cooperative). (B) Stochastic simulations of the non-cooperative model for parameters with an unstable (top) or stable (bottom) ON state (*che-1* expression), showing levels of *che-1* mRNA (blue) and protein (green). For parameters resulting in an unstable switch, stochastic fluctuations induce a spontaneous transition to the OFF state (no *che-1* expression). (C) Average ON state lifetimes calculated using Forward Flux Sampling (FFS) as function of CHE-1 dissociation rates from its own promoter (b_o or b_1, b_2) or and its target genes (b_t) for the non-cooperative and cooperative model. Stable ON state occurs for high *che-1* promoter occupancy by CHE-1 ($b_o < 1$ or $b_1, b_2 < 1$) and preferential affinity of CHE-1 for its own promoter compared to that of its target genes ($b_o \ll b_t$ or $b_2 \ll b_t$). (D) Average CHE-1 occupancy of the promoter of *che-1* (top) and other target genes (bottom), and average *che-1* mRNA level (middle) during spontaneous transitions from the ON to the OFF state, as sampled by FFS. Shown are transition paths for parameters with low (red, $b_o = 100 \text{ s}^{-1}$), medium (magenta, $b_o = 10 \text{ s}^{-1}$), and high (blue, $b_o = 1 \text{ s}^{-1}$) stability of the ON state, with $b_t = 10 \text{ s}^{-1}$. For simulations with a stable ON state, the *che-1* promoter remained fully occupied by CHE-1,

even as CHE-1 protein levels approached zero, in contrast to the occupancy of promoters of other CHE-1 target genes. **(E)** Simulations showing the impact of transient depletion of CHE-1 protein (green) on mRNA levels of *che-1* (blue) and a target gene (red). CHE-1 is depleted to 100 molecules/cell by a transient increase in degradation (b_c ; grey region). For parameters with an unstable ON state (top), both *che-1* and target gene mRNA levels fall rapidly, and do not recover when CHE-1 depletion ceases. For a stable ON state (bottom), expression of *che-1* is unaffected by CHE-1 depletion, leading to full recovery once CHE-1 depletion ends. **(F)** Average ON state lifetimes, calculated by FFS, during CHE-1 depletion to 100 molecules/cell. Parameter combinations with a stable ON state under normal conditions maintain *che-1* expression for hours or days under induced CHE-1 depletion.

The two models have 8 and 9 parameters, respectively, of which experimental data fixed 6 (see Methods). The production and degradation rates of *che-1* mRNA (f_m, b_m) and protein (f_c, b_c) were fully determined by the measured copy numbers and lifetimes. For the CHE-1 binding rate (f_o) we assumed the diffusion-limited rate, i.e. the highest physically possible binding rate. Based both on previous analysis (Etchberger et al. 2007) and our own, we examined model dynamics for ≈ 500 or 1000 CHE-1 target genes, i.e. either smaller or larger than the mean number of CHE-1 proteins (900 molecules/cell). The only free parameters were dissociation rates of CHE-1 from its own promoter (b_o or b_1, b_2) and from the other targets (b_T). We varied these between $0.1\text{-}100\text{ s}^{-1}$, corresponding to dissociation constants of $\sim 1\text{-}10^3\text{ nM}$ and consistent with values measured for transcription factors (Gebhardt et al. 2013; Jung et al. 2018). We simulated the models using the Gillespie algorithm (Gillespie 1977). In general, our stochastic simulations showed that molecular noise was sufficiently strong to induce spontaneous transitions from the ON to the OFF state on the timescale of hours or days, indicating that the measured copy numbers and lifetimes by themselves were not sufficient to generate stability against fluctuations. However, we also identified parameter combinations for which the CHE-1 switch remained in the ON state for at least a week (Fig. 4B).

For simulations with high switch stability, brute-force Gillespie simulations were too computationally demanding to directly measure the ON state lifetime. Instead, we used Forward Flux Sampling (FFS), a computational method to efficiently sample rare transition paths between states in multi-stable systems (Allen, Frenkel, and ten Wolde 2006). Using this approach, we observed parameter combinations with very high stability, i.e. lifetimes of many years, independent of the degree of cooperativity or number of target genes (Fig. 4C, S3A,B). In general, we observed increasing lifetimes for decreasing b_o or b_1, b_2 , i.e. higher occupancy of the *che-1* promoter by CHE-1. Moreover, the longest lifetimes were found when $b_o, b_2 \ll b_T$, i.e. when CHE-1 had a much higher affinity for its own promoter than for its other targets. In this regime, we observed average lifetimes of $> 1 \cdot 10^5$ years. Note, however, that despite the long lifetimes in this regime, spontaneous transitions to the OFF state are rapid and occur as a random Poisson process, with a transition

possible at any time, albeit with low probability. For such a Poisson process, ON state lifetimes of $> 2 \cdot 10^5$ years are required for the probability of spontaneous loss of ASE fate during their ~ 2 week lifetime to be less than 10^{-6} (see Methods), the frequency of spontaneous mutations per gene per generation in *C. elegans* (Anderson 1995).

3.3.4 Stability against stochastic fluctuations by preferential binding of CHE-1 to its own promoter

We found that high stability of the ON state required $b_o, b_2 \ll b_T$, i.e. CHE-1 binds its own promoter much more strongly than that of its other targets. An explanation for this emerged when we compared transition paths for spontaneous transitions to the OFF state, between parameter combinations that exhibited high ($> 1 \cdot 10^5$ years lifetime), medium (~ 12 days lifetime) and low (~ 5 hrs lifetime) stability (Fig. 4D, S3C). For parameters with low stability we found that, as CHE-1 protein levels fell during spontaneous transitions to the OFF state, both the average *che-1* mRNA number and the fraction of CHE-1 target promoters occupied by CHE-1 decreased, with very low occupancy even of the *che-1* promoter itself close to the end of the transition. In contrast, for parameters with high stability, we found that *che-1* promoter occupancy was high, and the *che-1* promoter remained bound by CHE-1 until the end of the transition, whereas CHE-1 binding was lost earlier on other promoters. These results suggested that high stability arises as a result of a strong preference for CHE-1 protein to bind to the *che-1* promoter, thereby making *che-1* expression insensitive to strong, stochastic decreases in CHE-1 level.

To test this idea, we ran simulations that included a transient depletion of CHE-1, implemented by a temporary increase in the CHE-1 protein degradation rate reducing CHE-1 to 100 molecules/cell (Fig. 4E). Indeed, we found that simulations with unstable switches, i.e. where $b_o, b_2 > b_T$ were highly sensitive to such depletions, with *che-1* mRNA rapidly falling to such low levels that CHE-1 protein levels and, hence, the ON state, were not recovered when CHE-1 depletion ceased. In contrast, even though the mRNA levels of target genes fell rapidly, simulations with highly stable switches maintained normal *che-1* mRNA levels for many days (Fig. 4F, S3D), allowing the system to successfully recover CHE-1 protein levels and the ON state if CHE-1 depletion was removed sufficiently rapidly. Finally, our simulations reproduced the stochastic nature of CHE-1 recovery observed experimentally (Fig. 1F, S3E), with lack of recovery the consequence of stochastic loss of all CHE-1 proteins during the period of induced depletion.

3.3.5 *In vivo* CHE-1 depletion decreases target gene but not *che-1* expression.

These simulation results were similar to our experimental observation that CHE-1::GFP levels fully recovered even after 24-48 hours of induced CHE-1 depletion (Fig. 1C-F, S1D,E). To test whether this reflected insensitivity of *che-1* expression to low CHE-1 protein levels (Fig. 4E,F), we used smFISH to compare the impact of auxin-mediated CHE-1::GFP::AID depletion at the mRNA levels of *che-1* and other target genes, focusing on *gcy-22* as the most highly expressed in our panel (Fig. 5A,B). Indeed, *gcy-22* mRNA levels were very low in most *che-1::GFP::AID* animals after 24 hrs on auxin. In striking contrast, *che-1* mRNA levels were close to wild-type. After 24 hrs without auxin, *gcy-22* mRNA levels had increased significantly (Fig. 5A,B), consistent with the recovery of CHE-1::GFP levels and chemotaxis to NaCl in these animals (Fig. 1D,E). Overall, these results were in full agreement with our model predictions but raised the question what properties of the *che-1* promoter were responsible for its resilience to CHE-1 depletion.

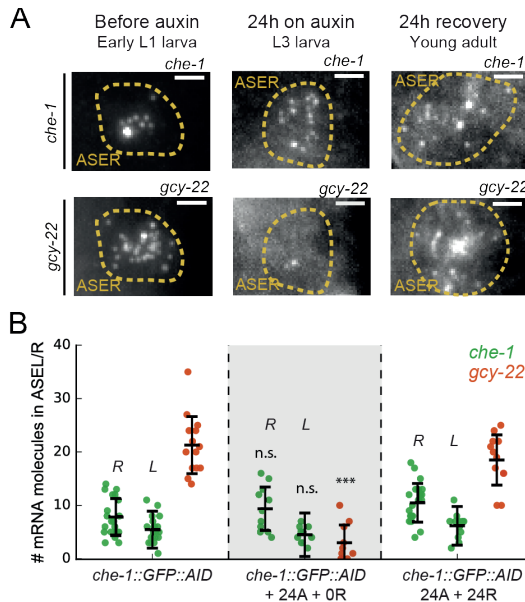


Figure 5. Maintenance of *che-1* expression during transient CHE-1 depletion. (A) *che-1* and *gcy-22* mRNA levels in ASER neurons, visualized by smFISH in *che-1::GFP::AID* animals, at different moments in a 24 hrs auxin and 24 hrs recovery treatment. Scale bar: 2 mm. (B) *che-1* (green, ASER/R) and *gcy-22* (red, ASER) mRNA levels quantified in *che-1::GFP::AID* animals at different times in a 24 hrs auxin and 24 hrs recovery treatment. Upon depletion of CHE-1 protein, *che-1* expression was not impacted, while *gcy-22* levels strongly decreased. *gcy-22* expression rose to wild-type levels after 24 hrs recovery off auxin. Error bars in B represent mean \pm SD. $n \geq 10$. *** $P < 0.001$.

3.3.6 Sequences flanking the *che-1* ASE motif are required for ***che-1*** expression during CHE-1 depletion

Previous studies identified a single ASE motif 242 bp upstream of the *che-1* ATG start codon as required for autoregulation of *che-1* expression (Etchberger et al. 2007; Leyva-Diaz and Hobert 2019). This ASE motif differs in 6 bp from the sequence in the *gcy-22* promoter, which might explain the divergent effects of CHE-1 depletion on *che-1* and *gcy-22* expression. To test this hypothesis, we used CRISPR/Cas9 in *che-1::GFP::AID* animals to replace the 12 bp ASE motif in the *che-1* promoter with that of *gcy-22*, and vice versa (Fig. 6A). The resulting mutant animals showed wild-type chemotaxis to NaCl and exhibited *che-1* and *gcy-22* mRNA levels similar to wild-type (Fig. 6B,F, S4A), indicating that replacing ASE motifs did not impact ASE specification and target gene expression. Moreover, when we depleted CHE-1 protein using auxin, *gcy-22* expression in $(ASE_{che-1})p::gcy-22$ animals almost completely vanished (Fig. 6F), as in animals with a wild-type *gcy-22* promoter. Overall, the ASE motif itself could not explain the observed differences in *che-1* and *gcy-22* expression. This agrees with previous results that showed a similar calculated affinity score of CHE-1 for the *che-1* and *gcy-22* ASE motif, despite sequence differences (Etchberger et al. 2009).

To examine if promoter regions other than the ASE motif were responsible for the differences in *che-1* and *gcy-22* expression, we replaced ASE motifs together with 59 bp flanks on either side (Fig. 6A). $(ASE_{gcy-22}+flanks)p::che-1$ animals exhibited a strong chemotaxis defect and lack of CHE-1::GFP expression in ASE neurons of larvae (Fig. 6B, S4B). At the pre-bean stage, $(ASE_{gcy-22}+flanks)p::che-1$ embryos showed high *che-1* expression, 28 ± 7 mRNA/cell, and CHE-1::GFP (Fig. 6C,D, S4B), indicating that *che-1* expression was initiated normally during ASE determination. However, *che-1* mRNA was absent in $(ASE_{gcy-22}+flanks)p::che-1$ embryos at later stages. This phenotype was also seen in the *che-1(p679)* loss of function mutant (Fig. S4C) and shows that this 130 bp *che-1* promoter region is important for maintenance of *che-1* expression and ASE fate. $(ASE_{che-1}+flanks)p::gcy-22$ animals showed wild-type chemotaxis to NaCl (Fig. 6B) and *gcy-22* mRNA levels (Fig. 6F). Yet, strikingly, upon CHE-1::GFP::AID depletion in $(ASE_{che-1}+flanks)p::gcy-22$ animals *gcy-22* mRNA levels remained high (Fig. 6F), indicating that, like *che-1*, *gcy-22* expression became resilient to CHE-1 depletion. Hence, the 130 bp *che-1* promoter fragment surrounding the ASE motif is responsible for maintaining expression during CHE-1 depletion.

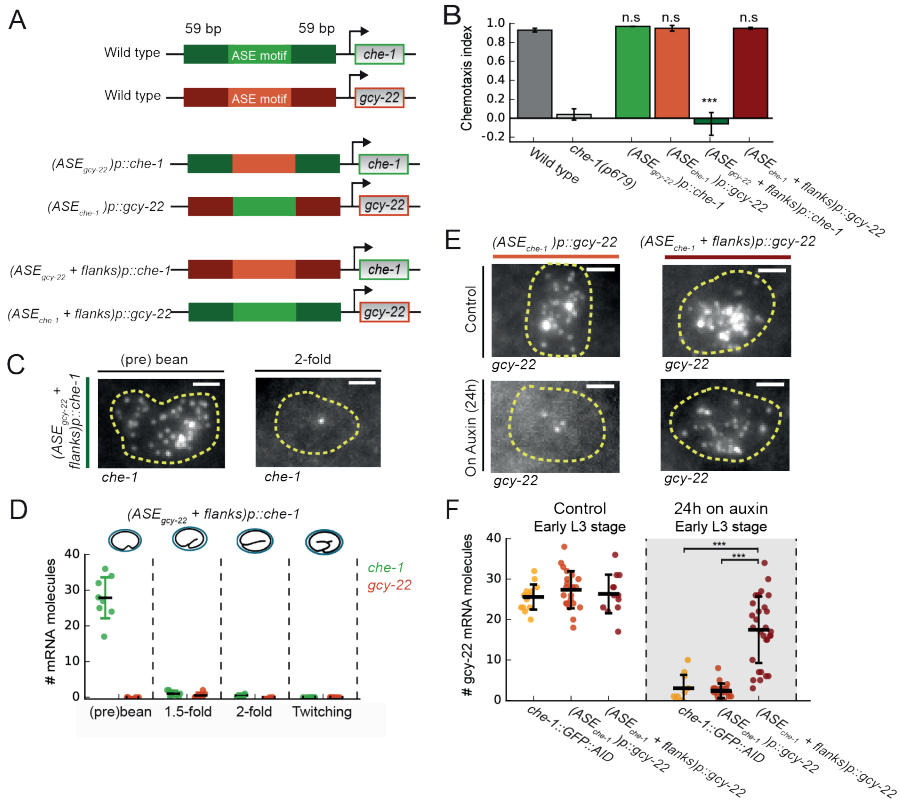


Figure 6. Region flanking CHE-1 binding site ensures resilience to CHE-1 depletion. (A) Overview of *che-1* promoter mutants generated in the *che-1::GFP::AID* background. We exchanged either the *che-1* (green) or *gcy-22* (red) ASE motif, which binds CHE-1, or a larger region that includes 59 bp flanks on either side. **(B)** Average chemotaxis index for response to 10 mM NaCl, of wild-type and *che-1(p679)* animals, and *che-1* promoter mutants. Exchange of ASE motifs between *che-1* and *gcy-22* promoters did not affect chemotaxis. Replacing the *che-1* ASE motif with flanks for that of *gcy-22* abolished chemotaxis to NaCl; the reverse had no effect. **(C)** *che-1* expression visualized by smFISH in ASE neurons of *(ASE_{gcy-22} + flanks)p::che-1* embryos. In 2-fold embryos, *che-1* mRNA levels were low. Scale bar: 1.5 mm. **(D)** *che-1* and *gcy-22* mRNA levels quantified in the ASE neurons of *(ASE_{gcy-22} + flanks)p::che-1* embryos. After initial high *che-1* expression at the time of ASE specification, *che-1* and *gcy-22* expression was almost absent, indicating a failure of ASE fate maintenance. **(E)** *gcy-22* expression under normal conditions or upon CHE-1 depletion by auxin, in L3 larvae of *che-1* promoter mutants. Scale bar: 1.5 mm. **(F)** Quantification of *gcy-22* mRNA levels upon auxin-induced CHE-1 depletion, in *gcy-22* promoter mutants. In *(ASE_{gcy-22})p::che-1* animals, *gcy-22* mRNA levels fell on auxin, as observed before. However, in *(ASE_{che-1} + flanks)p::gcy-22* animals treated with auxin *gcy-22* levels remained high. Thus, the region flanking the *che-1* ASE motif drives the maintenance of *che-1* expression during CHE-1 protein depletion. Error bars in B represent mean \pm S.E.M, D and F represent mean \pm SD. $n \geq 10$. *** $P < 0.001$.

3.3.7 Involvement of an *Otx*-related homeodomain binding site in maintaining ASE subtype

Within the 130 bp *che-1* promoter fragment required for resilient *che-1* expression, we identified a high scoring *Otx*-related homeodomain transcription factor (HD-TF) binding site, 29 bp downstream of the *che-1* ASE motif (Fig. 7A), with only ~60 CHE-1 targets exhibiting a binding site of similar score within 100 bp of their ASE motifs (Table S2). To test the function of this HD-TF binding site, we deleted it from the *che-1* promoter in the *che-1::GFP::AID* background. These (ΔHD)*p::che-1* animals showed an intermediate chemotaxis defect (Fig. 7D). To examine whether this defect reflected changes in *che-1* expression we scored CHE-1::GFP expression at different larval stages (Fig. 7B). CHE-1::GFP was expressed in all embryos, but was progressively lost over time, with CHE-1::GFP absent in more than half of the young adults (2/18 animals for ASEL and 9/20 for ASER), indicating a defect in maintenance of *che-1* expression, not in ASE determination. We then used time-lapse microscopy (Gritti et al. 2016) to monitor the dynamics of CHE-1::GFP expression directly in single (ΔHD)*p::che-1* larvae (Fig. S5A). Strikingly, CHE-1::GFP expression was lost in a rapid, stochastic manner (Fig. 7E, S5A), as would be expected for spontaneous, noise-driven transitions to the OFF state. Theoretical studies showed that the rate of such transitions increases dramatically with decreasing copy number of the key transcription factors involved (Warren and ten Wolde 2005). Indeed, (ΔHD)*p::che-1* animals showed lower CHE-1::GFP fluorescence, corresponding to 190 ± 70 CHE-1 proteins/cell (Fig. 7E, S5B) and decreased *che-1* mRNA levels, with 1 ± 1 mRNAs/cell in L1 larvae (Fig. 7C). In addition, CHE-1::GFP expression was lost more often and earlier in development in ASEL neurons (Fig. 7B,E), which have lower average *che-1* mRNA and protein copy numbers than ASER neurons (Fig. 2D,F). Altogether, these results point to homeodomain proteins, that bind the HD-TF binding site, as essential for long-term maintenance of *che-1* expression and thus ASE cell fate, likely by protecting the ON state against low CHE-1 copy number fluctuations.

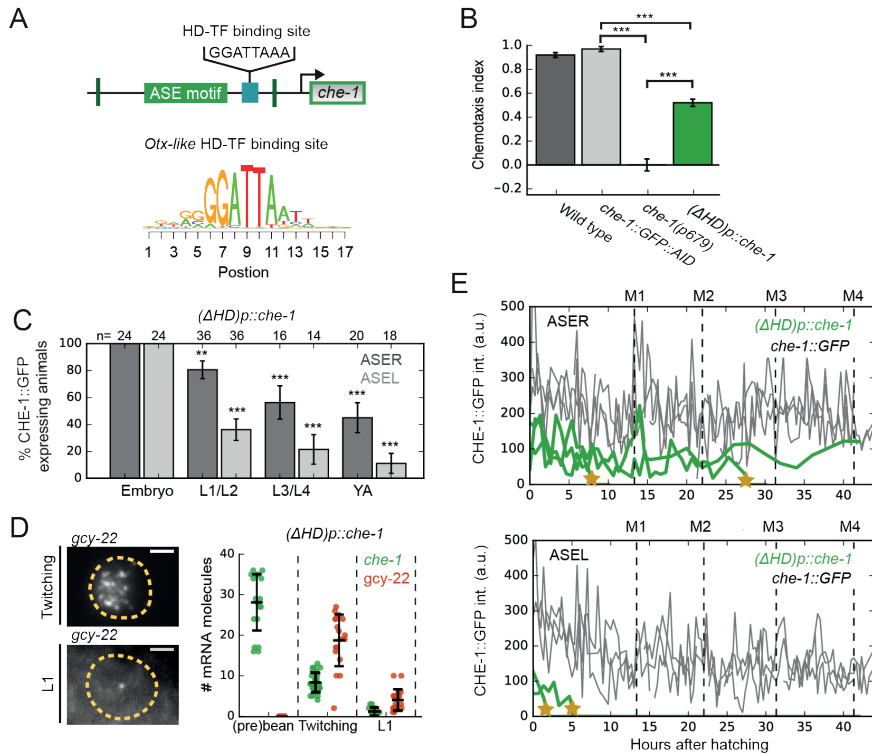


Figure 7. An Otx-related homeodomain transcription factor binding site involved in long-term maintenance of ASE cell fate. (A) Position of an Otx-related homeodomain transcription factor (HD-TF) binding site in the *che-1* promoter. Green lines indicate the positions of the 59 bp flanks surrounding the ASE motif. HD-TF binding site depicted as a sequence logo. The HD-TF binding site is deleted in $(\Delta HD)p::che-1$ animals. (B) Average chemotaxis index for response to 10 mM NaCl, of wild-type, *che-1::GFP::AID* and *che-1(p679)* animals, and $(\Delta HD)p::che-1$ mutant. Deleting the HD-TF binding site caused a decreased response to NaCl. (C) Fraction of $(\Delta HD)p::che-1$ animals expressing CHE-1::GFP in ASER (dark grey) and ASEL (light grey) at different developmental stages. CHE-1::GFP is progressively lost during development. (D) *che-1* and *gcy-22* mRNA levels in $(\Delta HD)p::che-1$ animals quantified by smFISH. Expression was similar to wild-type until late-stage, twitching embryos, but fell rapidly in newly-hatched L1 larvae. Scale bar: 2 mm. (E) CHE-1::GFP expression dynamics in single *che-1::GFP* (grey) and $(\Delta HD)p::che-1$ (green) animals during larval development. Approximate timing of molts is indicated M1-M4. CHE-1::GFP expression in $(\Delta HD)p::che-1$ animals was lost in a rapid and stochastic manner, at different times during development. Error bars represent S.E.M (B,D) or S.D. (C). $n \geq 10$. ** $P < 0.01$, *** $P < 0.001$.

3.4 Discussion

The terminal selector gene *che-1* controls subtype determination of the salt-sensing ASE neurons, by inducing the expression of hundreds of ASE-specific target genes while also inducing its own expression via autoregulation (Etchberger et al. 2007). A previous study showed that upon inhibition of positive CHE-1 autoregulation, transient *che-1* induction did not result in sustained *che-1* expression, with ASE cell fate lost shortly after its determination (Leyva-Diaz and Hobert 2019). However, this result left open whether this positive feedback loop by itself is sufficient to maintain *che-1* expression and ASE subtype for the animal's entire lifetime, or whether additional mechanisms, such as chromatin modifications, are required after initial determination to lock in cell identity in an irreversible manner. Here, we show that transient depletion of CHE-1 is sufficient to permanently lose ASE function, indicating that indeed *che-1* expression forms a reversible, bistable switch. This also raised the question how, in the presence of the inherent molecular fluctuations in the cell, this switch maintains its ON state (*che-1* expression) and prevent spontaneous transitions to the OFF state (no *che-1* expression).

Theoretical studies found that the stability of bistable genetic switches against fluctuations is enhanced by increasing the copy number and lifetime of the transcription factors that form the switch (Warren and ten Wolde 2005; Walczak, Onuchic, and Wolynes 2005; Mehta, Mukhopadhyay, and Wingreen 2008). However, our simulations using *in vivo* measured *che-1* mRNA and protein copy numbers and lifetimes showed that these values by themselves were not sufficient to generate a stable ON state. Instead, our simulations suggested a novel mechanism required to stabilize the ON state, which uses the reservoir of CHE-1 proteins bound to its many target sites to buffer *che-1* expression against fluctuations in CHE-1 level. Crucial to this target reservoir buffering mechanism is that CHE-1 exhibits strong preferential binding to its own promoter compared to its other targets. In this case, when CHE-1 protein levels drop, any CHE-1 protein that dissociates from a target gene promoter will immediately bind the *che-1* promoter, if it is found unoccupied. This ensures maintenance of *che-1* expression, at the expense of expression of its other targets, to bring CHE-1 protein levels back to the normal steady state level. Our experimental observations support this mechanism. Most importantly, upon transient CHE-1 depletion *in vivo*, expression of the CHE-1 target *gcy-22* rapidly and strongly decreased, yet *che-1* expression itself was hardly affected. We show that this resilience of *che-1* expression to CHE-1 depletion could be conferred onto other target genes by introducing a 130 bp fragment of the *che-1* promoter that surrounds the ASE motif bound by CHE-1, but not the ASE motif itself.

Target reservoir buffering relies on CHE-1 being preferentially recruited to

the *che-1* promoter. Our work suggests that homeodomain transcription factors are responsible for this, as deleting a HD-TF binding site close to the *che-1* ASE motif and specific to the *che-1* promoter, caused spontaneous transitions to the OFF state, and loss of cell identity, long after ASE subtype determination. This lack of stability was accompanied by lower *che-1* mRNA and protein levels, consistent with reduced recruitment of *che-1* to its own promoter. We speculate that factors binding the HD-TF binding site might act by increasing the residence time of CHE-1 at the *che-1* promoter. Our simulations showed that highly stable ON states require a relatively small, 10-100 fold increase in CHE-1 residence time at the *che-1* promoter compared to its other targets, within the range expected if CHE-1 and a HD-TF interact cooperatively with a weak interaction of a few $k_B T$. A similar interaction was postulated for the homeodomain protein ALR-1 in mechanosensory TRN neurons, where it restricts variability in expression of the terminal selector gene *mec-3*, by binding close to the MEC-3 binding site on its promoter (Topalidou, van Oudenaarden, and Chalfie 2011). The only protein with high predicted affinity to the HD-TF binding site in the *che-1* promoter, and expressed in ASE neurons, is the *Otx*-related homeodomain transcription factor CEH-36 (Lanjuin et al. 2003). However, a *ceh-36* loss-of-function mutant (Chang, Johnston, and Hobert 2003) did not reproduce the loss of ASE identity seen upon deletion of the homeodomain binding site (data not shown), indicating a role for other ASE-expressed homeodomain proteins, potentially in combination with CEH-36.

The observed resilience of *che-1* expression to CHE-1 depletion is crucial for stable maintenance of ASE identity. In both our models with and without cooperativity, this is achieved by a very low threshold for inducing *che-1* expression, with a single *che-1* mRNA potentially sufficient to induce the ON state. Indeed, the recovery of CHE-1::GFP expression after 24-48 hrs of CHE-1 depletion demonstrates that the ON state can be recovered from very low CHE-1 levels. However, this raises the question how stochastic, spontaneous induction of *che-1* expression and ASE fate is prevented in non-ASE cells. In most cells, spontaneous induction of the ON state is likely prevented by chromatin remodelling (Tursun et al. 2011; Patel and Hobert 2017). However, transient *che-1* expression induced long-term expression of CHE-1 targets in non-ASE head sensory neurons, suggesting that these cells are capable of inducing the ON state (Tursun et al. 2011). A potential mechanism to prevent ectopic induction of the ON state is provided by our observation that preferential binding of CHE-1 to its own promoter likely depends on homeodomain proteins. Cells lacking these proteins would have difficulty inducing and maintaining *che-1* expression. Consistent with this hypothesis, a large number of HD-TFs is expressed in the *C. elegans* nervous system in a highly neuron-specific manner (Hobert 2010). Our work suggests that cell-specific expression of co-factors of terminal selector genes might form a general mechanism to prevent spontaneous induction of these

genes in the wrong cells.

While bistability in genetic networks is recognized as an important mechanism to generate cell fate switches (Ferrell 2002), long-term cell fate maintenance is often assumed to require additional feedback mechanisms, for instance through histone and chromatin modifications, that make cell fate essentially irreversible. Here, we show that bistability through an autoregulatory feedback loop alone is sufficient for life-long maintenance of neuron identity in *C. elegans*, despite strong stochastic molecular fluctuations in the underlying genetic network. The mechanism responsible for this, target reservoir buffering, depends crucially on the single-input module topology of the network, with a terminal selector, CHE-1, inducing both its own expression and that of many other target genes. Single-input modules are highly prevalent network motifs for neuron subtype determination, both in *C. elegans* and higher organisms (Hobert and Kratsios 2019). In addition, cell differentiation in general is often controlled by a small number of master regulators that, directly or indirectly, induce both their own expression and that of many cell fate-specific target genes. Hence, we expect target reservoir buffering to play an important, general role in explaining stable long-term maintenance of cell fate in a broad array of systems.

Our work also underscores the link between biophysical parameters, such as molecular copy number and half-life, and stable cell fate maintenance. In mammals, stable maintenance of (neuronal) cell fate is even more challenging, given that the number of cells and their required lifetime is vastly larger compared to *C. elegans*. It will be important to determine if simply increasing copy number and half-life of a mammalian terminal selector is already sufficient, when coupled with target reservoir buffering, to explain the exceeding stability of cell fate in mammals. Our work now provides an experimental and theoretical framework to address this question.

3.5 Material and methods

C. elegans strains and handling

The following alleles were used in this study: *che-1*(*ot856[che-1::GFP]*) (kindly provided by Dylan Rahe from the Hobert lab) (Leyva-Diaz and Hobert, 2019), *gcy-5*(*ot835[gcy-5::SL2::mNeonGreen]*) II; *otTi6[hsp16-41p::che-1::2xFLAG]* X, *che-1*(*p679*), *ieSi57[Peft-3::TIR1::mRuby::unc-54 3'UTR, Cbr-unc-119(+)]* II. A complete list of strains used in this study, including transgenic strains generated by CRISPR/Cas9, is available upon request.

The wild-type strain used was the *C. elegans* variety Bristol, strain N2. All *C. elegans* strains were maintained on Nematode Growth Medium (NGM) plates, containing *E. coli* strain OP50 as a food source, at 20°C (Brenner, 1974), unless indicated otherwise. Worms were maintained according to standard protocol.

CRISPR/Cas9-mediated genome editing

Genome editing was performed according to protocol (Dokshin et al., 2018) and using ssODN repair templates with 35 bp homology arms (Paix et al., 2017). An AID tag was endogenously inserted at the C-terminus of GFP in a *che-1*(*ot856[che-1::GFP]*) background using guide g2 and a repair template containing the degon sequence (Zhang et al., 2015), generating *che-1*(*gj2089[che-1::GFP::AID]*). Subsequently, the *ieSi57* allele (Zhang et al., 2015) was introduced. All *che-1* or *gcy-22* promoter mutations were made in this *che-1*(*gj2089[che-1::GFP::AID]*;*ieSi57* background. (*ASE_{gcy-22}+flanks*)*p::che-1* was generated using a template containing the ASE motif of the *gcy-22* promoter and its 59 bp flanking regions, and guides g51 and g52. (*ASE_{che-1}+flanks*)*p::gcy-22* was generated using a template containing the ASE motif of *che-1* and its 59 bp flanks and guides g55 and g56. (*ASE_{gcy-22}*)*p::che-1* was generated using a template containing the ASE motif of *gcy-22* and guide g53. (*ASE_{che-1}*)*p::gcy-22* was generated using a template containing the ASE motif of *che-1* and guide g57. (*ΔHD*)*p::che-1* was generated using a template containing the 5 bp flanks of the HD motif from the promoter of *che-1* and guide g54. Guides and ssODNs used in this research are listed in Table S4, available upon request.

Single Molecule Fluorescence In Situ Hybridization (smFISH)

The oligonucleotides for the smFISH probe sets were designed with optimal GC content and specificity for the gene of interest using the Stellaris RNA FISH probe designer. The oligonucleotides were synthesized with a 3' amino C7 modification and purified by LGC Biosearch Technologies. Conjugation of the oligonucleotides with either Cy5 (GE Amersham) or Alexa594 (Invitrogen) was done as previously described (Lyubimova, 2013). Sequences of each probe set are listed in Table S3 (with exception of *gcy-14*) and are available upon request. The smFISH protocol was performed as previously described (Raj et al., 2008; Ji and van Oudenaarden 2012). Briefly, staged animals were washed from plates with M9 buffer and fixed in 4% formaldehyde in 1x PBS, gently rocking at room temperature (RT) for 40 minutes (young adults for 35 minutes). Fixation of embryos required a snap-freeze step to crack the eggshells by submerging embryos, after 15 minutes in fixation buffer, in liquid nitrogen, and thawing on ice for 20 minutes. After fixation, the animals were 2x washed with 1xPBS and resuspended in 70% ethanol overnight at 4°C. Ethanol was removed and animals were washed with 10% formamide and 2X SSC, as preparation for the hybridization. Animals were incubated with the smFISH probes overnight in the dark at 37°C in a hybridization solution (Stellaris) with added 10% formamide. The next day, animals were washed 2x with 10% formamide and 2X SSC each with an incubation step of 30 minutes at 37°C. The last wash step contains DAPI 5 µg/mL for nuclear staining. The wash buffer was removed, and animals were resuspended in 2X SSC and stored at 4°C until imaging. The 2X SSC was aspirated and animals were immersed in 100 µl GLOX buffer (0.4% glucose, 10 mM Tris-HCl, pH 8.0, 2X SSC) together with 1 µl Catalase (Sigma-Aldrich) and 1 µl glucose oxidase (Sigma-Aldrich) (3.7 mg/mL) to prevent bleaching during imaging.

Microscopy images of smFISH samples were acquired with a Nikon Ti-E inverted fluorescence microscope, equipped with a 100X plan-apochromat oil-immersion objective and an Andor Ikon-M CCD camera controlled by µManager software (Edelstein et al., 2014). smFISH analysis was performed with custom Python software, based on a previously described method (Raj et al., 2008). Briefly, we first convolved the smFISH images with a Gaussian filter. Next, candidate spots were selected via manual thresholding, and partially overlapping spots were separated via finding 3D regional intensity maxima. We used the spatial localization of *gcy-22* or *gcy-14* mRNA molecules which are highly expressed ASE-specific genes, to estimate the cell boundaries of the ASE neurons. The coverage of the ASE cell bodies with *gcy-22* or *gcy-14* mRNA molecules agreed with GFP markers (*gcy-5p::GFP* or *gcy-7p::GFP*) that

marked the cell body.

CHE-1 protein quantification

Staged *che-1::GFP* knock-in animals were bathed in 72 nM and 48 nM eGFP recombinant protein (Bio-connect) with 0.25 mM Levamisole (Sigma-Aldrich) in M9 buffer in a glass chambered cover glass systems (IBL baustoff), which were coated with 0.5 mg/ml kappa-capsein in the buffer MRB80 (80 mM Pipes, 4 mM MgCl_2 , 1 mM EGTA, pH 6.8 with KOH) to prevent binding of eGFP to the cover glass and chamber walls. Images of bathed *che-1::GFP* animals in eGFP solution were acquired with a Nikon Eclipse Ti inverted microscope, equipped with a Nikon C1 confocal scan head, a 100 mW Argon ion laser (488nm, Coherent), and a S Fluor 40×1.3 NA and an Apo TIRF 100×1.49 NA objective. Calibration of eGFP with *che-1::GFP* animals was repeated 2 times at different days with 72 nM and 48 nM eGFP concentrations, of which we took the average calibration measurements. For ease of measuring, the CHE-1::GFP signal of animals was measured with the exact same microscope and software settings, except placing the animals submersed in 0.25 mM Levamisole (Sigma-Aldrich) in M9 buffer on agar pads with the same cover glass thickness on top. The ASE neuron closest to the cover glass was imaged in larvae to circumvent tissue scattering. Embryos were followed in time (at 22°C) and imaged every 20 minutes with the exact same microscope and software settings, from bean stage until twitching started. For both larvae and embryos, the slices focused at the approximate middle of the ASE neuron nuclei were used for quantifying the CHE-1::GFP signal. The volumes of the nuclei were calculated by measuring the radii of the nuclei in x, y and z direction from the CHE-1::GFP signal with the assumption that the nucleus shape resembles a ellipsoid, using the following equation: $V = \frac{4}{3} \pi xyz$.

FRAP

To estimate the protein degradation rate of CHE-1::GFP, we used Fluorescence Recovery After Photobleaching (FRAP). Animals were immobilized using Polybead microspheres (Polyscience) in M9 buffer on agarose pads covered with a cover glass. To prevent dehydration of animals, the coverslip was sealed with VALAP (vaseline, lanolin and paraffin, ratio 1:1:1). Animals were monitored at several time points during the experiment if they were still alive by checking very subtle movement and/or pumping behaviour. The GFP signal in the ASE neurons of the animals was bleached until approximately 20% of the initial signal was left. After bleaching, the GFP signal was measured every 20 or 30 minutes until the signal had recovered. Images were taken with the same microscope as in the CHE-1 protein quantification section. We measured for each time point the average GFP intensity in the ASE

neurons and subtracted the background intensity measured nearby the ASE neurons. The degradation rate was calculated from the initial slope of the growth curve using the following exponential model: $R(t) = \left(\frac{f}{b}\right) + \left(x_0 - \left(\frac{f}{b}\right)\right)e^{-bt}$, where x_0 is the initial fluorescent intensity at the start of the recovery curve right after bleaching. b and f represent the CHE-1 protein degradation and production rate, which are fitted on the individual measured recovery curves, to obtain the average CHE-1 protein degradation rate.

che-1 mRNA stability

We induced *che-1* mRNA overexpression with the *otTi6* [*hsp16-41p::che-1::2xFLAG*] X inducible heat shock strain, and we quantified with smFISH *che-1* mRNA in the ASER neurons in L2 larvae over time ($t_i=0, \dots, 4$) ~17 min apart until recovery. We determined the relative amount of *che-1* mRNAs from the start of the measurement, $n(t_i) = \frac{N(t_i) - N_{control}}{N_{HS} - N_{control}}$, where $N_{control}$ is the calculated average amount of *che-1* mRNAs when there is no *che-1* mRNA overexpression, N_{HS} the calculated average amount of *che-1* mRNAs at the first measured time point right after heat shock induction, and $n(t)$ is the amount of *che-1* mRNAs at the three remaining time points. An exponential degradation curve e^{-at} was fitted to the experimentally determined values $n(t_i)$, to obtain the approximate *che-1* mRNA degradation rate.

Mathematical Model of the CHE-1 switch

The cooperative mathematical model assumes that CHE-1 protein (C) has to bind as a dimer on the *che-1* promoter to induce *che-1* mRNA expression. The binding and unbinding of the two CHE-1 proteins at the *che-1* promoter (O) is separated into two events, first one CHE-1 binds with binding rate f_o and unbinds with unbinding rate b_1 from the *che-1* promoter (OC). When the first CHE-1 protein is bound to the *che-1* promoter, the second CHE-1 protein binds with the same binding rate f_o next to the first CHE-1, forming a dimer on the *che-1* promoter (OC2), and unbinds with the unbinding rate b_2 . CHE-1 proteins bind as a monomer on the target gene promoters ($O_T C$), with the unbinding rate b_T , and with same the binding rate f_o as on the *che-1* promoter. *che-1* mRNA (M) is transcribed with the production rate f_M only when two CHE-1 proteins are bound to the *che-1* promoter, and *che-1* mRNA is translated into CHE-1 protein with the protein production rate f_C . Both *che-1* mRNA and CHE-1 protein are degraded with the degradation rates, b_M and b_C respectively. This leads to the following differential equations:

$$\frac{dOC}{dt} = f_o(O^* - (OC + OC2))C - b_1OC$$

$$\frac{dOC2}{dt} = f_oOC \cdot C - b_2OC2$$

$$\frac{dO_T C}{dt} = f_o(O_T^* - O_T C)C - b_T O_T \cdot C$$

$$\frac{dM}{dt} = f_M OC2 - b_M M$$

$$\frac{dC}{dt} = -f_o(O^* - (OC + OC2))C - f_oOC \cdot C - f_o(O_T^* - O_T C)C + b_1OC + b_2OC2 + b_T O_T \cdot C + f_c M - b_c C$$

Where O^* is the total number of *che-1* promoters and where O_T^* is the total number of target genes.

In the non-cooperative model, CHE-1 binds (f_o) and unbinds (b_o) as a monomer on the *che-1* promoter (OC) and induces *che-1* expression as a monomer. The other reactions in the model are the same as in the cooperative model. This leads to the following differential equations:

$$\frac{dOC}{dt} = f_o(O^* - OC)C - b_o OC$$

$$\frac{dO_T C}{dt} = f_o(O_T^* - O_T C)C - b_T O_T \cdot C$$

$$\frac{dM}{dt} = f_M OC - b_M M$$

$$\frac{dC}{dt} = -f_o(O^* - OC)C - f_o(O_T^* - O_T C)C + b_o OC + b_T O_T \cdot C + f_c M - b_c C$$

Bistability

In the cooperative model CHE-1 proteins bind as a dimer at the *che-1* promoter to induce *che-1* expression. The binding of two CHE-1 proteins in the system, implies non-linear behaviour, giving rise to bistability. We have in the cooperative model three fixed points, 2 stable and 1 unstable fixed point. The 2 stable fixed points represent the so-called “ON state” (high CHE-1) and the “OFF state” (low CHE-1) of the CHE-1 switch. When the switch is in the OFF state, it has to cross the unstable point to reach the ON state. In contrast, the non-cooperative model, in which CHE-1 proteins bind as a monomer at the *che-1* promoter to induce *che-1* expression, has only two fixed points. The first fixed point represents the ON state, in which CHE-1 protein levels are high. The second fixed point is a half-stable point, when there is no CHE-1 protein, the switch is OFF. However, the introduction of, for example, only 1 *che-1* mRNA in the system would be sufficient to turn the switch from the OFF state to the ON state.

Simulations

To study the short-term lifetime (<1 weeks) of the cooperative and non-cooperative CHE-1 switch, we performed stochastic Gillespie simulations on both models (Gillespie 1977). We used a custom written python script to simulate the reactions involved in both CHE-1 switch models. All reactions describing the differential equations from the cooperative and non-cooperative model, are summarized in Fig. 3A. Parameters remain unchanged during simulations (with exception for transient depletion simulations) and species are initiated in the ON state.

To study ON state lifetimes of the CHE-1 switch exceeding 1 weeks, in the cooperative and non-cooperative CHE-1 switch model, we used Forward Flux Sampling (Allen et al., 2006; Allen et al., 2009), a computational method which allowed us to estimate lifetimes of CHE-1 switches. It was not necessary to integrate pruning into the algorithm since this would not result in improvement in computational efficiency. The FFS algorithm was initiated with the same initial conditions as the Gillespie simulations. Interfaces of the FFS algorithm were chosen according to the variance of CHE-1 protein in the ON state, to generate a 5-10% chance of CHE-1 protein trajectories crossing the first interface (with an exception for very unstable switch, where CHE-1 protein simulations immediately run to the OFF state). The typical step size of the interfaces was 20 and the number of interfaces was between 20 and 35.

Parameters

We can divide the parameters in the following groups: 1.) experimentally determined parameters, 2.) parameters that we could approximate, 3.) unconstrained parameters and 4.) parameters that we could calculate with help of the other parameters. First, the parameters which were experimentally determined. We based the *che-1* mRNA degradation rate (b_m) on direct measurement of *che-1* mRNA degradation that we quantified in the ASER neuron (Fig 2A-B). We used an approximation of *che-1* mRNA lifetime of 20 minutes. The amount of CHE-1 protein C was set to 900 molecules (average of ASER and ASEL at L4/YA stage) based on the CHE-1 protein quantification experiments, and the amount of *che-1* mRNAs (M) was set to 7 molecules (average of ASER and ASEL at L4 stage) based on the smFISH experiments in wild-type animals.

The binding rate of CHE-1 at its own promoter f_o and the target gene promoters f_o is approximated by the diffusion limited binding rate. The diffusion limited binding rate of CHE-1 on promoter sites was calculated using the following diffusion equation: $f_D = 4\pi\sigma D$, where reaction cross section $\sigma = 1 \cdot 10^{-2} \mu m$, i.e. the size of the promoter binding site, and the diffusion coefficient constant is $D = 1 \mu m^2 s^{-1}$. To obtain the diffusion coefficient that we can apply in our simulations, which includes information about the volume of the nucleus, the diffusion coefficient f_D was divided with the average nucleus volume (V_c) of $4 \mu m^3$ in ASER and ASEL, resulting in a diffusion limited binding rate of $f_o = \frac{f_D}{V_c} = 0.03 s^{-1}$.

To approximate the number of target sites where the CHE-1 protein can bind, we used the crude approximation depicted from the study of Etchberger et al. (2007), in which they found 596 genes represented with at least three times as many tags in the ASE versus AFD SAGE library, which is expected to contain almost no “false positives”. We acquired the total amount of ASE motifs (minimum score of 98%) in each of the promoter regions (<1000 bp upstream of the start site) with a custom written R script, using the TFBSTools/JASPAR2018 packages, resulting in a total of 425 ASE motifs. However, the data set lacked, for example, housekeeping genes expressed in the ASE neurons that could be under control of *che-1*. The sci-RNA-seq dataset from the study of Cao et al. (2017), provides information on genes expected to be expressed in the ASE neurons, including ASE non-specific genes (Cao et al., 2017). A total of 1400-1500 genes are expected to be expressed in the ASEL and ASER (score >100, removal of “false-positives”), resulting in ~1000 ASE motifs on which CHE-1 could potentially bind. In the simulations we used either 500 or 1000 target sites.

The unbinding rates of CHE-1 from its own promoter and the target gene promoters,

b_p, b_d, b_T are the only unconstrained parameters, and are ranged between 0.1 and 100 sec^{-1} in the FFS simulations. The CHE-1 protein degradation rate b_C , the CHE-1 production rate f_C and the *che-1* mRNA production rate f_M are dependent on the unbound fraction of CHE-1 protein. To calculate the three unknown rates b_C, f_C, f_M first the CHE-1 protein degradation rate b_C was set at the experimentally measured parameter, and used to calculate the *che-1* mRNA production rate f_M and CHE-1 protein production rate f_C . The CHE-1 production rate is influenced by the amount of CHE-1 bound at the target gene promoters, hence why the CHE-1 protein production and degradation had to be fitted to the experimentally measured CHE-1 protein FRAP curve. The fitting was done by introducing CHE-1::GFP species, bleached or unbleached, in the existing model in order to reproduce in simulation the FRAP experiments. The bleached CHE-1::GFP proteins could also bind and unbind the promoters and induce *che-1* expression, but the bleached CHE-1::GFP proteins could not be produced, only degraded.

In the table below, we summarize all parameter values. In this table, all parameter values are defined for the *Unstable CHE-1 switch* (4B), and the other model parameter values are only given when they deviate from those used in the *Unstable CHE-1 switch* (4B).

Chapter 3

Parameter	Description	Value
Unstable swith (4b)	Unstable CHE-1 switch, average lifetime ~10 days	
C	Number of CHE-1 protein molecules	900
M	Number of <i>che-1</i> mRNA molecules	7
O_S	Number of <i>che-1</i> promoters	1
$O_{T,S}$	Number of target genes	500
f_O	Binding rate of CHE-1 on <i>che-1</i> or target gene promoters	0.03 sec^{-1}
b_O	Unbinding rate of CHE-1 from <i>che-1</i> promoter	100 sec^{-1}
b_T	Unbinding rate of CHE-1 from target gene promoters	100 sec^{-1}
f_M	<i>che-1</i> mRNA production rate	0.0302 sec^{-1}
b_M	<i>che-1</i> mRNA degradation rate	0.00083 sec^{-1}
f_C	CHE-1 protein production rate	0.0274 sec^{-1}
b_c	CHE-1 protein degradation rate	0.00024 sec^{-1}
Stable swith (4b)	Highly stable switch, slower unbinding rate of CHE-1 from its own promoter	
b_O	Unbinding rate of CHE-1 from <i>che-1</i> promoter	0.1 sec^{-1}
f_M	<i>che-1</i> mRNA production rate	0.0059 sec^{-1}
f_c	CHE-1 protein production rate	0.0261 sec^{-1}
b_c	CHE-1 protein degradation rate	0.00023 sec^{-1}
Stable switch (4E)	Highly stable switch, depleted to 100 CHE-1	
b_O	Unbinding rate of CHE-1 from <i>che-1</i> promoter	0.1 sec^{-1}
f_M	<i>che-1</i> mRNA production rate	0.0059 sec^{-1}
b_c	CHE-1 protein degradation rate	0.0019 sec^{-1}
f_C	CHE-1 protein production rate	0.024 sec^{-1}
b_p	Target mRNA degradation rate	0.0008 sec^{-1}
f_p	Target mRNA production rate	0.09 sec^{-1}
Unstable switch (4E)	Unstable switch, depleted to 100 CHE-1	
b_O	Unbinding rate of CHE-1 from <i>che-1</i> promoter	10 sec^{-1}
b_T	Unbinding rate of CHE-1 from target gene promoters	10 sec^{-1}
b_c	CHE-1 protein degradation rate	0.00076 sec^{-1}
f_C	CHE-1 protein production rate	0.026 sec^{-1}

The parameters for the depletion simulations panel (Fig 4E) include an extra promoter (PO) with production and degradation rates of (PM) mRNA, b_p, f_p to simulate the production of target mRNA induced by CHE-1. The binding and unbinding rates of the extra promoter are the same as for the target genes, f_C, b_T .

Estimation of required ON state lifetime.

If spontaneous switches from the ON to the OFF state occur as a Poisson process with rate r , then the probability of a switching event occurring at time t is given by the exponential distribution $p(t) = r \exp(-rt)$. The fraction of animals in which a switching event occurs before time $t=T$ is given by the cumulative distribution $P(t) = 1 - \exp(-rt)$. This fraction is smaller than a value ϕ if $r < -\ln(1 - \phi)/T$. For small fractions $\phi \approx 0$, this can be approximated as $r < \phi / T$. If a switching event can occur only in 1 out of 10^6 animals, $\phi = 1 \cdot 10^{-6}$, within the average animal lifetime of $T \approx 2$ weeks, then the required life time of the ON state is given by $1/r > 2.3 \cdot 10^5$ years.

Auxin induced protein degradation

The Auxin Inducible Degron (AID) System was employed as previously described (Zhang et al., 2015). Animals were initially grown on NGM plates with OP50 without IAA (Sigma-Aldrich). To induce degradation of CHE-1, L1 staged animals were transferred to NGM plates with OP50 and 1 mM IAA in EtOH at 20 °C.

For the CHE-1::GFP::AID measurements, the duration of auxin exposure and recovery were varied in each treatment. The animals were transferred every other day to new NGM plates with 1 mM IAA or recovery NGM plates without IAA to prevent mixing of generations. Each treatment contained a control group of *che-1::GFP::AID* animals never exposed to auxin. For imaging, the animals were placed on an 5% agarose pad submerged in 0.25 mM Levamisole (Sigma-Aldrich) in M9 with a cover glass on top. We checked for CHE-1::GFP::AID in the ASE neurons with a wide field microscope. Older animals showed stronger autofluorescence in the head, giving rise to nuclei-like structures which could be confused with ASE neurons. The GFP (FITC) channel was combined with the TRITC channel to correct for the autofluorescence (Teuscher and Ewald, 2018). Due to tissue scattering and decrease in CHE-1::GFP signal in old animals, we found the ASE neuron closest to the cover glass was most reliable. The control group always showed CHE-1::GFP::AID in the ASE neurons: even though CHE-1::GFP::AID levels decreased with age, even 10 days old animals had well identifiable ASE neurons.

NaCl chemotaxis assay

The quadrant assay used to assess chemotaxis to NaCl was adapted from Wicks et al. (2000) and Jansen et al. (2002). Briefly, two diagonally opposite quadrants of a sectional petri dish (Star Dish, Phoenix Biomedical) were filled with 13.5 ml buffered agar (1.7% Agar, 5 mM K_2HPO_4/KH_2PO_4 pH 6, 1 mM $CaCl_2$ and 1 mM $MgSO_4$) containing NaCl and two diagonally opposite quadrants with 13.5 ml buffered agar without NaCl. Immediately before the assay, the plastic dividers between

the quadrants were covered with a thin layer of agar. Age synchronized *C. elegans* populations, cultured at 25°C for ~66 hours, were washed 3 times for 5 min. with CTX buffer (5 mM K_2HPO_4/KH_2PO_4 pH 6, 1 mM $CaCl_2$ and 1 mM $MgSO_4$). Approximately 100 animals were placed in the middle of a sectional dish. After 10 min., animals on each quadrant were counted and a chemotaxis index (CI) was calculated for each plate ($CI = (\# \text{ animals on NaCl} - \# \text{ animals not on NaCl}) / \text{total} \# \text{ animals}$). To determine the CI of a strain, 2 assays per day were performed on at least 2 different days.

To assess the effect of CHE-1 depletion on chemotaxis, animals were bleached and cultured for 24 hours on NGM plates without IAA at 20°C. After 24 hours animals were transferred to NGM plates containing 1 mM IAA. To remove eggs and larvae, animals were thoroughly rinsed using CTX buffer and a 30 μm pluriStrainer (pluriSelect) and transferred to a fresh NGM plate, with or without 1 mM IAA, starting at 96 hours into the experiment and repeated every 24 hours until the end of the experiment. After the experimental treatment duration, the chemotaxis index (CI) was determined. Subsequently, recovery was started by transferring animals to NGM plates without IAA and the CI was determined 24 hours and 48 hours thereafter. At each timepoint, the CI was determined in similarly aged *che-1(p679)* and untreated *che-1::GFP::AID* animals.

Time lapse

Time-lapse imaging was performed as previously described (Gritti et al., 2016). Briefly, micro-chambers are made out of polyacrylamide hydrogel, made from a 10% dilution of 29:1 acrylamide/bis-acrylamide (Sigma-Aldrich) was mixed with 0.1% ammonium persulfate (Sigma-Aldrich) and 0.01% TEMED (Sigma-Aldrich) as polymerization initiators. For the time-lapse experiments the chambers were 240x240x20 μm , these dimensions were sufficient to contain enough OP50 bacteria to sustain development until animals started laying eggs.

We used a Nikon Ti-E inverted microscope with a 40x objective in all experiments. The microscope has a Hamamatsu Orca Flash 4.0 v2 camera set at full frame and full speed. The camera chip is 13x13 mm and contains 4Mp. We used 488 nm lasers (Coherent OBIS-LS 488-100) for fluorescence excitation. We used a high fluorescent signal of 100 mW with an exposure time of 10 ms, since the fluorescent signal of CHE-1::GFP is relatively low. To keep track of the molting cycle as indication of the age of the animals, we used bright field imaging, which contained a red LED (CoolLED, 630nm). Time-lapse images were acquired every 20 minutes without detectable phototoxicity effects.

Images were analyzed with custom written time-lapse software, and with ImageJ. Briefly, first we corrected the raw images for experimental aberrations with flat and dark field images acquired at the end of the experiment. For quantification purposes, we computed the average fluorescence of each ASE neuron via drawing a region of interest around each nucleus and we corrected the average intensity by subtracting the background level close to the ASE neuron.

Quantification and statistical analysis

Image analysis of CHE-1::GFP quantification and intensity measurements was performed with the ImageJ distribution Fiji (Schindelin et al. 2012). For the quantification data shown in graphs of all figures, the dots represent individual values, the boxplots without box represent the mean and the standard deviation.

Statistical analyses were performed either using R software, version 3.6.0, or with Python 3.5 Package SciPy. Comparisons of the chemotaxis indexes were performed using a one-way ANOVA, followed by a pairwise t-test with Holm correction. Significance between control versus conditions in smFISH data were performed using one-way ANOVA, followed by a Tukey multiple comparison test. Significance of % expression of CHE-1::GFP in ASE neurons between treatment groups (on auxin) and control group (no auxin) was performed using Fisher exact test.

3.6 Appendix

3.6.1 Acknowledgements

Some strains were provided by the CGC, which is funded by NIH Office of Research Infrastructure Programs (P40 OD010440), and the Mitani laboratory through the National Bio-Resource Project of the MEXT, Japan. We thank Dylan Rahe (Hobert lab) for providing the GFP-tagged che-1(ot856) allele. This work is part of the research program of the Foundation for Fundamental Research on Matter (FOM), which is financially supported by the Netherlands Organization for Scientific Research (NWO).

3.6.2 Author contributions

J.J.H.T, S.N.B., G.J. and J.S.Z conceived the study. J.J.H.T performed smFISH, time-lapse and (confocal) microscopy expression experiments and data analysis. S.N.B. generated strains, performed behavioural assays and data analysis. J.J.H.T and J.S.Z. performed mathematical modelling. J.J.H.T, S.N.B., G.J. and J.S.Z. wrote the manuscript.

3.6.3 Supplementary figures

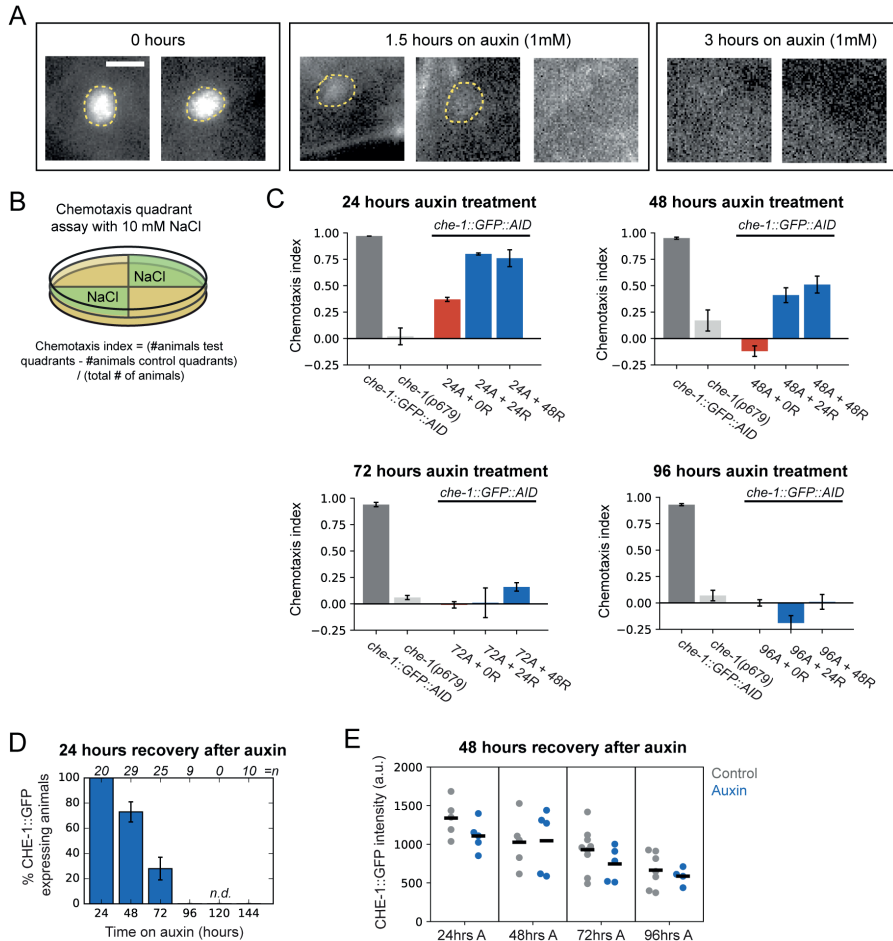


Figure S1. CHE-1 depletion leads to loss of ASE cell fate. (A). Representative images of *CHE-1::GFP* in animals after different durations of auxin-induced CHE-1 depletion. Dashed circles highlight visible *CHE-1::GFP* fluorescence. For animals after 1.5 hours on auxin, *CHE-1::GFP* fluorescence was much reduced, when visible, while for animals 3 hours on auxin no *CHE-1::GFP* was visible and the fluorescence signal at the expected location of the ASE neuron is shown instead. Overall, 15/15 control animals showed visible *CHE-1::GFP*, compared to 4/10 and 0/7 animals after 1.5 and 3 hours exposure to auxin, respectively, indicating that auxin-mediated CHE-1 depletion was rapid. Scalebar: 3 μ m (B) Schematic of quadrant chemotaxis assay and chemotaxis index (CI) calculation. (C) Chemotaxis index for wild-type, *che-1(p679)* and *che-1::GFP::AID* animals exposed to auxin for 24-96 hrs after a recovery period of 0, 24 or 48 hrs. (D) Percentage of animals that recovered *CHE-1::GFP* expression after auxin treatment of increasing length with a recovery period of 24 hours. (E) *CHE-1::GFP* fluorescence intensity in animals that recovered *CHE-1::GFP* after 24-96 hours of auxin-induced CHE-1 degradation, compared to control animals of the same

age that were not exposed to auxin. In animals that recovered CHE-1::GFP after transient CHE-1 depletion, fluorescence levels returned to similar values as control animals, indicating that CHE-1 recovery was complete.

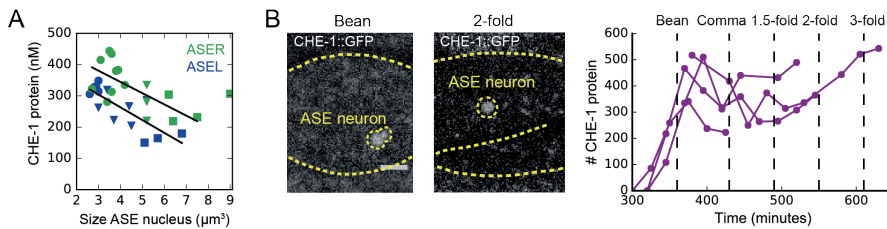


Figure S2. CHE-1 copy numbers in larvae and embryos. (A) CHE-1 protein concentration in ASER (green) and ASEL (blue) at different stages of post-embryonic development (L1/L2: circles, L3/L4: triangle, young adult: squares). CHE-1 protein concentration decreases with the age of the animals. (B) Images of *CHE-1::GFP* embryos for quantification of CHE-1::GFP in ASE neurons with confocal microscopy (left). Number of CHE-1 protein molecules over time in embryos, starting at the bean stage until twitching started (at 22 °C) (right). Embryos showed similar levels of CHE-1 protein as the L1 larvae.

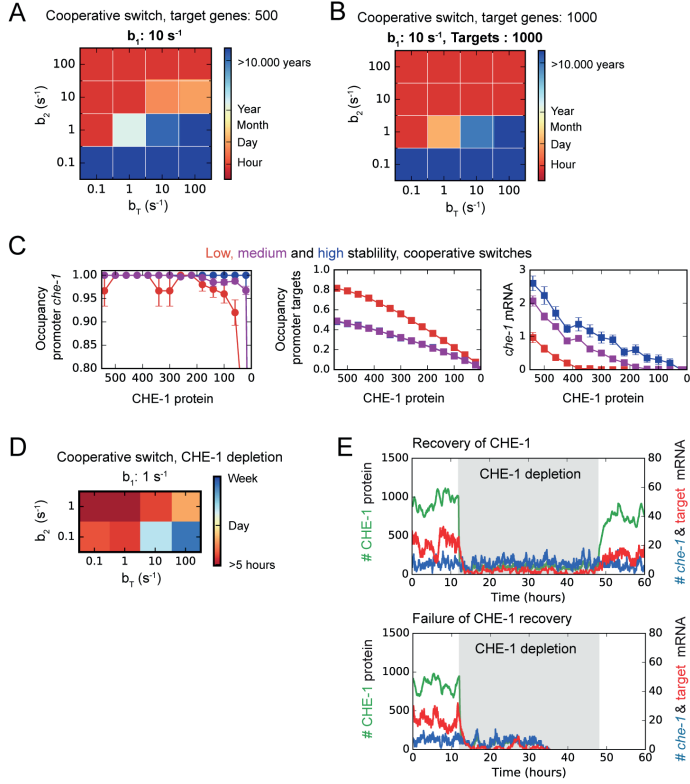


Figure S3. Dependence of ON state stability on number of targets and cooperativity. (A,B) Average ON state lifetimes calculated using Forward Flux Sampling (FFS) as function of the dissociation rates of CHE-1 from its own promoter (b_2, b_1) and target genes (b_r) promoters, for the cooperative model with $O_r^* = 500$ (A) or 1000 (B) CHE-1 targets. ON state lifetimes are lower for 1000 targets, but long lifetimes of many years are still found. $b_1=10 \text{ sec}^{-1}$ and b_1, b_r varied between $0.1 - 100 \text{ sec}^{-1}$. (C) Average CHE-1 occupancy of the promoter of *che-1* (left) and other target genes (middle), and average *che-1* mRNA level (right) during spontaneous transitions from the ON to the OFF state, as sampled by FFS, for the cooperative model. Shown are transition paths for parameters with low (red, $b_2 = 100 \text{ s}^{-1}, b_r = 10 \text{ s}^{-1}$), medium (magenta, $b_2 = 10 \text{ s}^{-1}, b_1 = 100 \text{ s}^{-1}$), and high (blue, $b_2 = 1 \text{ s}^{-1}, b_1 = 100 \text{ s}^{-1}$) stability of the ON state, with $b_1 = 10 \text{ s}^{-1}$. (D) Average ON state lifetimes calculated using Forward Flux Sampling (FFS) of the cooperative model during depletion of CHE-1 protein, as function of the dissociation rates of CHE-1 from its own promoter (b_2) and its target genes (b_r), where $b_1 = 10 \text{ sec}^{-1}$. (E) Examples highlighting that upon transient CHE-1 depletion (grey) some simulations recover CHE-1 levels (upper panel), while others lose CHE-1 permanently (lower panel). Parameter simulations and strength of CHE-1 depletion conditions are the same as in Fig. 4E in the main text, but duration of depletion is increased here from 24 to 48 hours.

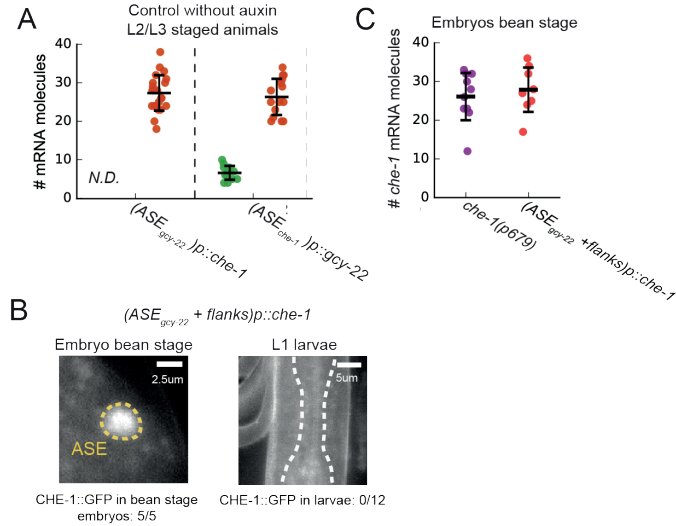


Figure S4. Controls for promoter region mutants. (A) Quantification of *gcy-22* (red) and *che-1* (green) mRNA levels in *(ASE_{gcy-22})p::che-1* and *(ASE_{che-1})p::gcy-22* L2/L3 animals without auxin revealed wild type mRNA levels for both genes. (B) CHE-1::GFP expression in *(ASE_{gcy-22} + flanks)p::che-1* bean stage embryos and L1 larvae. Embryos showed CHE-1::GFP expression (5/5 animals), whereas L1 larvae no longer expressed CHE-1::GFP (0/12 animals). (C) *(ASE_{gcy-22} + flanks)p::che-1* embryos showed similar levels of *che-1* mRNA during the bean stage as the *che-1(p679)* mutant.

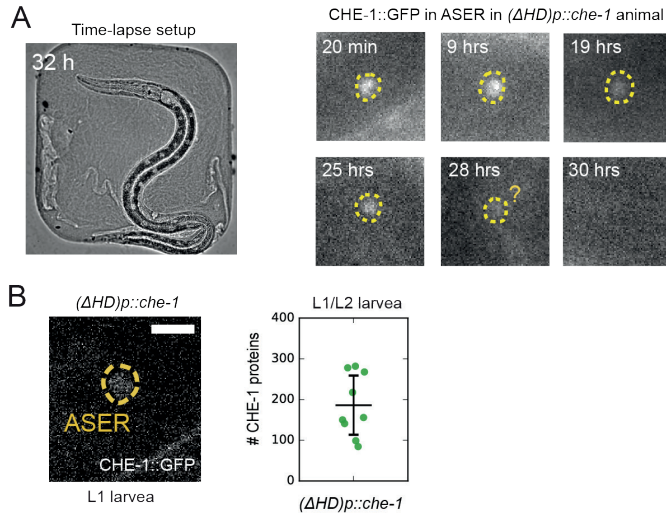


Figure S5. Time lapse imaging and quantification of CHE-1::GFP in $(\Delta HD)p::che-1$ animals. (A) Long-term time-lapse microscopy of a single $(\Delta HD)p::che-1$ animal, using microchambers to constrain larvae to the field of view (left). CHE-1::GFP expression in ASER disappeared rapidly in the L3 larval stage, 30 hours after hatching (right). (B) Image of CHE-1::GFP in $(\Delta HD)p::che-1$ L1 larva (left). Quantification of CHE-1::GFP protein in $(\Delta HD)p::che-1$ L1/L2 staged animals (right).

3.7 References

- Acar, M., Becskei, A., and van Oudenaarden, A. (2005). Enhancement of cellular memory by reducing stochastic transitions. *Nature* 435, 228-232.
- Allen, R.J., Frenkel, D., and ten Wolde, P.R. (2006). Forward flux sampling-type schemes for simulating rare events: efficiency analysis. *J Chem Phys* 124, 194111.
- Allen, R.J., Valeriani, C., and Rein Ten Wolde, P. (2009). Forward flux sampling for rare event simulations. *J Phys Condens Matter* 21, 463102.
- Alon, U. (2007). Network motifs: theory and experimental approaches. *Nat Rev Genet* 8, 450-461.
- Anderson, P. (1995). Mutagenesis. *Methods Cell Biol* 48, 31-58.
- Brenner, S. (1974). The genetics of *Caenorhabditis elegans*. *Genetics* 77, 71-94.
- Cao, J., Packer, J.S., Ramani, V., Cusanovich, D.A., Huynh, C., Daza, R., Qiu, X., Lee, C., Furlan, S.N., Steemers, F.J., *et al.* (2017). Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* 357, 661-667.
- Chang, S., Johnston, R.J., Jr., and Hobert, O. (2003). A transcriptional regulatory cascade that controls left/right asymmetry in chemosensory neurons of *C. elegans*. *Genes Dev* 17, 2123-2137.
- Coburn, C.M., and Bargmann, C.I. (1996). A putative cyclic nucleotide-gated channel is required for sensory development and function in *C. elegans*. *Neuron* 17, 695-706.
- Deneris, E.S., and Hobert, O. (2014). Maintenance of postmitotic neuronal cell identity. *Nat Neurosci* 17, 899-907.
- Dhondt, I., Petyuk, V.A., Bauer, S., Brewer, H.M., Smith, R.D., Depuydt, G., and Braeckman, B.P. (2017). Changes of Protein Turnover in Aging *Caenorhabditis elegans*. *Mol Cell Proteomics* 16, 1621-1633.
- Dokshin, G.A., Ghanta, K.S., Piscopo, K.M., and Mello, C.C. (2018). Robust Genome Editing with Short Single-Stranded and Long, Partially Single-Stranded DNA Donors in *Caenorhabditis elegans*. *Genetics* 210, 781-787.
- Edelstein, A.D., Tsuchida, M.A., Amodaj, N., Pinkard, H., Vale, R.D., and Stuurman, N. (2014). Advanced methods of microscope control using muManager software. *J Biol Methods* 1.
- Etchberger, J.F., Flowers, E.B., Poole, R.J., Bashllari, E., and Hobert, O. (2009). Cis-regulatory mechanisms of left/right asymmetric neuron-subtype specification in *C. elegans*. *Development* 136, 147-160.
- Etchberger, J.F., Lorch, A., Sleumer, M.C., Zapf, R., Jones, S.J., Marra, M.A., Holt, R.A., Moerman, D.G., and Hobert, O. (2007). The molecular signature and cis-regulatory architecture of a *C. elegans* gustatory neuron. *Genes Dev* 21, 1653-1674.
- Ferrell, J.E., Jr. (2002). Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. *Curr Opin Cell Biol* 14, 140-148.
- Fornasiero, E.F., Mandad, S., Wildhagen, H., Alevra, M., Rammner, B., Keihani, S., Opazo, F., Urban, I., Ischebeck, T., Sakib, M.S., *et al.* (2018). Precisely measured protein lifetimes in the mouse brain reveal differences across tissues and subcellular fractions. *Nat Commun* 9, 4230.
- Gebhardt, J.C., Suter, D.M., Roy, R., Zhao, Z.W., Chapman, A.R., Basu, S., Maniatis, T., and Xie, X.S. (2013). Single-molecule imaging of transcription factor binding to DNA in live mammalian cells. *Nat Methods* 10, 421-426.
- Gillespie, D.T.J.T.j.o.p.c. (1977). Exact stochastic simulation of coupled chemical reactions. *81*, 2340-2361.
- Gregor, T., Tank, D.W., Wieschaus, E.F., and Bialek, W. (2007). Probing the limits to

- positional information. *Cell* 130, 153-164.
- Gritti, N., Kienle, S., Filina, O., and van Zon, J.S. (2016). Long-term time-lapse microscopy of *C. elegans* post-embryonic development. *Nat Commun* 7, 12500.
- Hobert, O. (2008). Regulatory logic of neuronal diversity: terminal selector genes and selector motifs. *Proc Natl Acad Sci U S A* 105, 20067-20071.
- Hobert, O. (2010). Neurogenesis in the nematode *Caenorhabditis elegans*. *WormBook*, 1-24.
- Hobert, O. (2016). Terminal Selectors of Neuronal Identity. *Curr Top Dev Biol* 116, 455-475.
- Hobert, O., and Kratsios, P. (2019). Neuronal identity control by terminal selectors in worms, flies, and chordates. *Curr Opin Neurobiol* 56, 97-105.
- Hsiao, H.Y., Jukam, D., Johnston, R., and Desplan, C. (2013). The neuronal transcription factor *erect* wing regulates specification and maintenance of *Drosophila* R8 photoreceptor subtypes. *Dev Biol* 381, 482-490.
- Jansen, G., Weinkove, D., and Plasterk, R.H. (2002). The G-protein gamma subunit *gpc-1* of the nematode *C.elegans* is involved in taste adaptation. *EMBO J* 21, 986-994.
- Ji, N., and van Oudenaarden, A. (2012). Single molecule fluorescent in situ hybridization (smFISH) of *C. elegans* worms and embryos. *WormBook*, 1-16.
- Jung, C., Bandilla, P., von Reutern, M., Schnepf, M., Rieder, S., Unnerstall, U., and Gaul, U. (2018). True equilibrium measurement of transcription factor-DNA binding affinities using automated polarization microscopy. *Nat Commun* 9, 1605.
- Lanjuin, A., VanHoven, M.K., Bargmann, C.I., Thompson, J.K., and Sengupta, P. (2003). *Otx/otd* homeobox genes specify distinct sensory neuron identities in *C. elegans*. *Dev Cell* 5, 621-633.
- Leblond, C.P., and Walker, B.E. (1956). Renewal of cell populations. *Physiol Rev* 36, 255-276.
- Leyva-Diaz, E., and Hobert, O. (2019). Transcription factor autoregulation is required for acquisition and maintenance of neuronal identity. *Development* 146.
- Mehta, P., Mukhopadhyay, R., and Wingreen, N.S. (2008). Exponential sensitivity of noise-driven switching in genetic networks. *Phys Biol* 5, 026005.
- Ming, G.L., and Song, H. (2005). Adult neurogenesis in the mammalian central nervous system. *Annu Rev Neurosci* 28, 223-250.
- Ninkovic, J., Pinto, L., Petricca, S., Lepier, A., Sun, J., Rieger, M.A., Schroeder, T., Cvekl, A., Favor, J., and Gotz, M. (2010). The transcription factor *Pax6* regulates survival of dopaminergic olfactory bulb neurons via crystallin α A. *Neuron* 68, 682-694.
- Orlando, V. (2003). Polycomb, epigenomes, and control of cell identity. *Cell* 112, 599-606.
- Ortiz, C.O., Faumont, S., Takayama, J., Ahmed, H.K., Goldsmith, A.D., Pocock, R., McCormick, K.E., Kunimoto, H., Iino, Y., Lockery, S., *et al.* (2009). Lateralized gustatory behavior of *C. elegans* is controlled by specific receptor-type guanylyl cyclases. *Curr Biol* 19, 996-1004.
- Ozbudak, E.M., Thattai, M., Lim, H.N., Shraiman, B.I., and Van Oudenaarden, A. (2004). Multistability in the lactose utilization network of *Escherichia coli*. *Nature* 427, 737-740.
- Paix, A., Folkmann, A., and Seydoux, G. (2017). Precision genome editing using CRISPR-Cas9 and linear repair templates in *C. elegans*. *Methods* 121-122, 86-93.
- Patel, T., and Hobert, O. (2017). Coordinated control of terminal differentiation and restriction of cellular plasticity. *Elife* 6.
- Raj, A., van den Bogaard, P., Rifkin, S.A., van Oudenaarden, A., and Tyagi, S. (2008).

- Imaging individual mRNA molecules using multiple singly labeled probes. *Nat Methods* 5, 877-879.
- Ringrose, L., and Paro, R. (2007). Polycomb/Trithorax response elements and epigenetic memory of cell identity. *Development* 134, 223-232.
- Sarin, S., Antonio, C., Tursun, B., and Hobert, O. (2009). The *C. elegans* Tailless/TLX transcription factor nhr-67 controls neuronal identity and left/right asymmetric fate diversification. *Development* 136, 2933-2944.
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., *et al.* (2012). Fiji: an open-source platform for biological-image analysis. *Nat Methods* 9, 676-682.
- Schwanhauser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. *Nature* 473, 337-342.
- Serrano-Saiz, E., Leyva-Diaz, E., De La Cruz, E., and Hobert, O. (2018). BRN3-type POU Homeobox Genes Maintain the Identity of Mature Postmitotic Neurons in Nematodes and Mice. *Curr Biol* 28, 2813-2823 e2812.
- Sharova, L.V., Sharov, A.A., Nedorezov, T., Piao, Y., Shaik, N., and Ko, M.S. (2009). Database for mRNA half-life of 19 977 genes obtained by DNA microarray analysis of pluripotent and differentiating mouse embryonic stem cells. *DNA Res* 16, 45-58.
- Suel, G.M., Garcia-Ojalvo, J., Liberman, L.M., and Elowitz, M.B. (2006). An excitable gene regulatory circuit induces transient cellular differentiation. *Nature* 440, 545-550.
- Teuscher, A.C., and Ewald, C.Y. (2018). Overcoming Autofluorescence to Assess GFP Expression During Normal Physiology and Aging in *Caenorhabditis elegans*. *Bio Protoc* 8.
- Topalidou, I., van Oudenaarden, A., and Chalfie, M. (2011). *Caenorhabditis elegans* *aristaless/Arx* gene *alr-1* restricts variable gene expression. *Proc Natl Acad Sci U S A* 108, 4063-4068.
- Tursun, B., Patel, T., Kratsios, P., and Hobert, O. (2011). Direct conversion of *C. elegans* germ cells into specific neuron types. *Science* 331, 304-308.
- Walczak, A.M., Onuchic, J.N., and Wolynes, P.G. (2005). Absolute rate theories of epigenetic stability. *Proc Natl Acad Sci U S A* 102, 18926-18931.
- Warren, P.B., and ten Wolde, P.R. (2005). Chemical models of genetic toggle switches. *J Phys Chem B* 109, 6812-6823.
- Wicks, S.R., de Vries, C.J., van Luenen, H.G., and Plasterk, R.H. (2000). CHE-3, a cytosolic dynein heavy chain, is required for sensory cilia structure and function in *Caenorhabditis elegans*. *Dev Biol* 221, 295-307.
- Zhang, L., Ward, J.D., Cheng, Z., and Dernburg, A.F. (2015). The auxin-inducible degradation (AID) system enables versatile conditional protein depletion in *C. elegans*. *Development* 142, 4374-4384.

Chapter 4

Validation of pharynx and neuron related clusters found in RNA-sequencing data with novel density-based clustering approach

Joleen J. H. Traets, Dimphy van Boerdonk, Steffen Werner, Tom .S. Shimizu and
Jeroen S. van Zon

This chapter is a follow-up study of our previous publication:
Steffen Werner, W Mathijs Rozemuller, Annabel Ebbing, Anna Alemany, Joleen
Traets, Jeroen S. van Zon, Alexander van Oudenaarden, Hendrik C. Korswagen,
Greg J. Stephens, Thomas S. Shimizu. Functional modules from variable genes:
Leveraging percolation to analyze noisy, high-dimensional data. bioRxiv
2020.06.10.143743; doi: <https://doi.org/10.1101/2020.06.10.143743>

4.1 Abstract

Gene expression is inherently variable. Using a novel density-based clustering approach developed by Werner et al., we identified clusters of genes that exhibited correlated variability in expression level, in a single-animal RNA-sequencing data set of *C. elegans* nematode worms synchronized at the same developmental stage. Even though some clusters could be explained at least partly by developmentally-controlled changes in gene expression, the origin of two clusters remained unclear. The first cluster consists out of genes related to neurons, and the second cluster is related to pharynx structure. Moreover, the clusters not only showed correlated gene expression between gene pairs within each cluster individually, but also between both clusters. Here, we aimed to validate both clusters and to provide insight into their nature and function. We quantified expression levels of representative genes for each cluster with single molecule FISH (smFISH), to test for correlations in gene expression between the gene pairs. The selected genes from the pharynx and neuron cluster were almost solely expressed the pharynx region and nerve ring region, respectively. We were able to validate correlated variability in expression in the pharynx cluster, linking it to oscillations in gene expression that occur up until the

young adult stage. We explored a possible link between ancestral starvation history and gene expression in the neuron cluster, but found that this did not enhance correlations within or between the clusters. However, the TGF- β ligand *daf-7* had a modulation in expression levels in animals whose ancestors were starved. Moreover, direct starvation induced expression levels of *daf-7*, only in animals at the same developmental stage that expression of the pharynx cluster genes peak, increasing correlation between the two clusters. This result could explain the correlation between the clusters observed by RNA-sequencing, as short-term starvation likely occurred in the RNA-sequencing protocol used in our experiments.

4.2 Introduction

Even within unperturbed isogenic populations of multi-cellular organisms, often substantial phenotypic variation exists between individuals (Kirkwood et al. 2005; Gartner 1990; Wong, Gottesman, and Petronis 2005). Drivers of phenotypic variation between individuals potentially include information on environmental history that can be passed on in a transgenerational manner (Webster et al. 2018; Moore, Kaletsky, and Murphy 2019; Rando 2012). Also, maternal age at time of conception has also emerged as a driver of variation between individuals (Perez et al. 2017). Another important underlying driver of the phenotypic variation are the inherently stochastic processes within cells (Raj and van Oudenaarden 2008; Kaern et al. 2005).

The information carried by the variability in gene expression within unperturbed isogenic populations has been exploited to increase understanding of biological processes, for example to perform gene network prediction (Munsky, Neuert, and van Oudenaarden 2012), aging (Kirkwood et al. 2005) and study transgenerational memory (Perez et al. 2017). Many studies exploiting variation in gene expression in order to find gene regulatory interactions use RNA-sequencing methods, allowing gene expression profiling on a genome-wide scale. Traditionally, RNA-sequencing analysis has focused on differential analysis of gene expression between manipulated and control conditions (Anders et al. 2013; Han et al. 2015). However, advances in single- cell or organism techniques now allow for exploiting the variations in unperturbed populations, e.g. in classifying cell types based on differences in transcriptome within a heterogeneous population (Grun and van Oudenaarden 2015; Wagner, Regev, and Yosef 2016; Luecken and Theis 2019). In this Chapter, we will build on results from density-based clustering approach developed recently by us (Werner et al. 2020), that addresses the last challenge; identifying the underlying regulatory relationships among genes by variability in gene expression across individual samples from unperturbed populations.

The novel density-based clustering approach focusses on identifying functional modules (Werner et al. 2020), and is motivated by the theory of percolation on random graphs (Penrose and Pisztor 1996; Penrose, Penrose, and Press 2003; Dall and Christensen 2002; Newman 2018). The algorithm exploits the generic behaviour of random geometric networks close to the percolation critical point, from which a null model for the noise is devised. This null model of the noise in turn provides a basis for identifying statistically significant branches within the cluster hierarchy. The approach was able to identify functional gene clusters from single cell expression data on fission yeast cells (Saint et al. 2019). Applying this analysis on data for multicellular organisms posed a more complex challenge, since RNA sequencing of individual animals results in expression data pooled from

a heterogenous population of the different cell types found in the animal's body. The approach was applied to a data set comprised of N=36 genetically identical *C. elegans* nematodes synchronized at young adult age. The animals developed under identical conditions. In total, we identified 13 functional gene clusters associated with specific biological functions and spatial expression (Werner et al. 2020). For example, we identified a cluster consisting almost solely out of genes related to sperm production. These genes are known to exhibit co-regulated expression dynamics in time, and our identification of this cluster shows that it is possible to extract such information based on variability in developmental stage between animals in the RNA-sequencing experiment.

It is impossible to synchronize the animals at the exact same age for the RNA-sequencing experiments, meaning that animals in our RNA-sequencing experiments exhibited slight variations in developmental age. Indeed, we could explain the existence of some clusters at least partly by developmentally-controlled changes in gene expression (Werner et al. 2020). However, the emergence of two specific clusters, labelled Cluster 1 and 2, remained unclear. Cluster 1 mainly consist out of genes related to the pharynx structure and are solely expressed in the pharynx region. Cluster 2 mainly consists out of genes related to neurons, such as neuropeptides, and are exclusively expressed in the head and tail region. Both clusters do not have a correlation with the body length of the individual animals, which would indicate expression dynamics as function of developmental stage. Most surprising, we found that variation in gene expression among individual animals was even correlated between these two clusters, even as a clear biological relation between these clusters was lacking.

In this study, we attempted to validate the existence of Clusters 1 and 2, and examined the origin of the correlated variation in expression observed between the clusters in the RNA sequencing data. We used single molecule FISH (smFISH) (Raj et al. 2008) to quantify expression of selected genes from both clusters, and characterize their spatial expression patterns in staged young adult animals within the same age ranges as in the RNA-sequencing experiment. The smFISH method makes it possible to visualize individual mRNA molecules in fixed tissue, thereby offering a highly sensitive quantification of gene expression levels with cellular resolution. In particular, we used smFISH to simultaneously visualize expression of pairs of genes in individual animals, to check for significant correlations between gene pairs, between those that belong to the same cluster and those from separate clusters.

We found that the selected pharynx cluster genes were all expressed exclusively in the pharynx region, and the selected neuron cluster genes were almost solely

spatially expressed in the nerve ring region, and there mostly in neuronal cells. We were able to validate at least one of the clusters, as the pharynx cluster genes showed significant correlations between each other, likely due to oscillations in gene expression that occur up until the young adult stage. Ancestral starvation history did not induce gene expression of the cluster genes or enhance the correlations within or between the cluster, as was suggested by previously published RNA-sequencing experiments (Webster et al. 2018). However, we did observe modulation of expression levels of the neuron cluster gene *daf-7* in animals as function of ancestral starvation history. Moreover, we found that upon direct starvation, expression levels of *daf-7* increased stronger in animals right after the fourth molt, at the same developmental stage that expression of the pharynx cluster genes increases. This result suggested that short-term starvation in the RNA-sequencing protocol used in our experiments could have induced the observed correlation between both clusters. Unfortunately, the neuron clusters could not be validated, since two out of the three selected genes form the neuron cluster either lacked strong evidence for belonging to the neuron cluster or could not be quantified with smFISH.

4.3 Results

4.3.1 Choice of genes representative for two clusters found in RNA-sequencing data with novel density-based clustering approach

Briefly, by employing the novel density-based clustering approach developed by Werner *et al.* (Werner et al. 2020) to RNA-sequencing data of synchronized hermaphrodites at the young adult stage (40 hours after hatching on HB101, N=34), we discovered several novel clusters (Fig. 1A). A cluster is defined as a group of genes showing co-variance in gene expression levels among the RNA-sequencing samples. Two clusters revealed interesting features: Cluster 1 (neuron cluster) contained almost only neuron-related genes and is explicitly expressed in the head and tail of the animal (N=56), while Cluster 2 (pharynx cluster) only consist of pharynx-related genes and is explicitly expressed in the pharynx region (N=55) (Fig. 1A-B). Moreover, the expression level averaged over all genes in each cluster varies between individual animals in a manner that is correlated between the two clusters (Fig. 1B). We performed additional expression analysis on a published RNA-sequencing data set on gene expression dynamics during *C. elegans* development (Hendriks et al. 2014), revealing that the pharynx cluster consists mainly out of genes that show oscillatory expression, while the expression levels of the neuron cluster genes remain constant (Fig. 2D-E). The oscillations could explain the co-variance in expression of the pharynx cluster genes. However, it remained uncertain if there are oscillations after the fourth molt during young adult stage since the RNA-sequencing data set by Hendriks *et al.* reaches only until the start of the fourth molt.

Hence, if the neuron and pharynx clusters are real and what caused their presence in the RNA-sequencing data set, remained elusive.

To validate both the neuron and pharynx cluster, we used single molecule FISH (smFISH) (Raj et al. 2008) to quantify expression of selected genes from both clusters in animals of similar age as in the RNA-sequencing data. With the smFISH method it is possible to visualize individual mRNA molecules, offering an independent, high-resolution quantification method for gene expression. We chose 3 genes from each cluster, 6 genes in total, to examine their expression levels and patterns. The choice of candidate genes was dependent on three requirements: 1) the gene should have a high score for the chance of being in the core of the cluster, meaning that the gene is representative for the variations in gene expression of most genes from the cluster (Werner et al. 2020), 2) the spliced mRNA of the gene must at least 25 probes of ~20bp in order to achieve a reliable smFISH signal, 3) the probes must have a GC content ideally around 45% (Ji and van Oudenaarden 2012). Satisfying all three requirements decreased our list of possible candidates severely, especially in the case of the neuron cluster, where most genes encode neuropeptides and hence are too short for smFISH to work reliably.

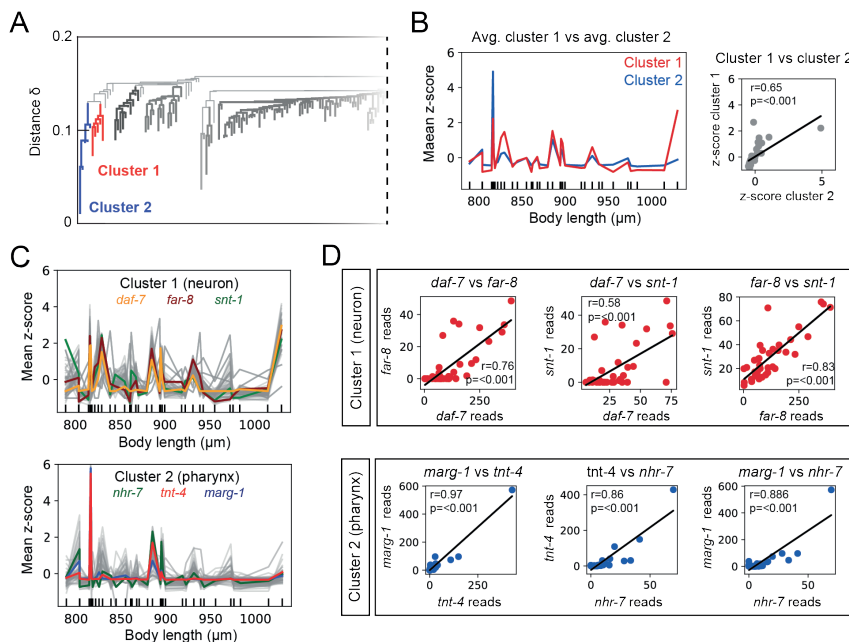


Figure 1. Selected genes from the neuron and pharynx cluster revealed with novel density-based clustering approach. (A) The novel density-based clustering approach by Werner *et al.*

(Werner *et al.*, 2020) on single-worm RNA-sequencing data from young adult animals (with $N = 34$) identified 13 clusters (grey shaded and coloured branches) with tissue-specific annotation. Red and blue clusters represent the neuron and pharynx cluster respectively. **(B)** Co-variance of Cluster 1 (neurons, in red) and Cluster 2 (pharynx, in blue) in the RNA-sequencing data between the individual animals with their respective body length (represented by vertical lines on x-axis). The correlation between both clusters is significant ($p < 0.001$, right panel). **(C)** The neuron cluster (top) and the pharynx cluster (bottom) together with the panel of genes selected for each cluster for smFISH experiments. **(D)** Correlations between the genes in the panel for each cluster reveal a significant correlation for each of the pairs (neuron: top, pharynx: bottom). Significance of correlations was measured with Spearman's rank correlation. In (B-C), black insets on the x-axis represent the individual animals.

For the neuron cluster, we chose the following genes: *daf-7*, *far-8* and *snt-1*, and for the pharynx cluster, we chose *marg-1*, *nhr-7* and *tnt-4*. For both clusters, the chosen genes co-vary in the RNA-sequencing data with the other genes in the same cluster (Fig. 1C), and the genes were also significantly correlated ($p < 0.001$) between one another (Fig. 1D). Moreover, the chosen genes representing the pharynx cluster also show significant correlation between one another without the obvious outlier in the data ($p < 0.05$, Fig. S1). This panel of genes meets the requirements stated previously, with each gene covered by 30-48 smFISH probes, sufficient for smFISH to achieve a reliable signal. Hence, all 6 genes should be representative for the features of the two cluster that were found in the RNA-sequencing data.

4.3.2 Highly specific expression pattern and biological functions of cluster genes

The biological functions of the genes in the neuron and pharynx clusters have not been studied extensively yet. To acquire more knowledge on relevant biological functions of both clusters, we used tissue, phenotype and GO-term enrichment analysis, and performed analysis using other publicly available data sets.

Expression pattern enrichment of the neuron cluster genes showed a strong enrichment for neuron and neuron-related cell structures (Fig. 2A). Moreover, TOMO-sequencing data in which individual animals are sequenced in separate slices from the tip of the head to the tail, showed that the neuron cluster genes are mainly expressed in the head and the tail of the animals where most neurons in *C. elegans* are located (Ebbing *et al.* 2018) (Fig. 2C-D). The phenotype enrichment for the neuron cluster genes contained mostly terms related to neuron functioning, in particular on the activity of body movement of the animals (Fig. 2A). The most significant phenotypes are related to increased body movement of the animals and suggest a relation with increased foraging behavior, e.g. forward motion increased, pausing decreased and foraging hyperactive. The GO term enrichment analysis of

the neuron cluster showed mostly terms related to signaling, cell communication and reaction to responses, which is in line with our previous finding on neuron-related activity.

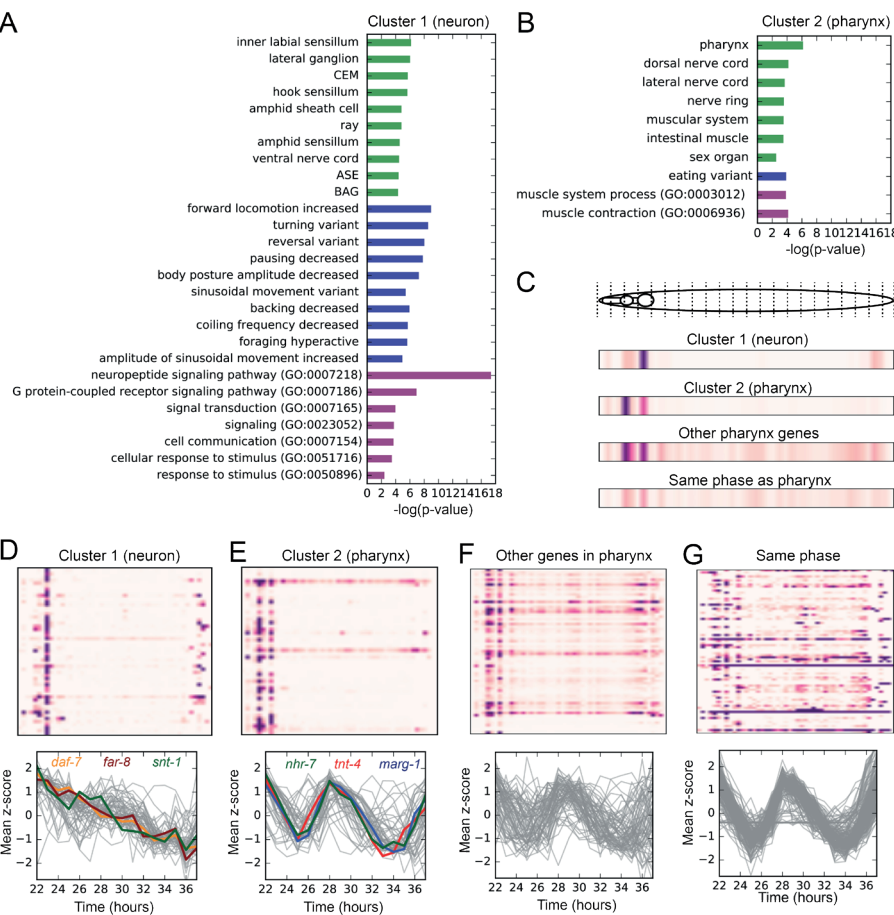


Figure 2: Highly specific expression of cluster genes in related tissue according to biological functions. (A-B) Tissue (green), phenotype (blue) and GO-term enrichment (purple) on the neuron (A) and pharynx (B) cluster genes. The cluster genes are enriched against a standard reference set of genes in order to determine significance of resulting terms (see Methods). (C) TOMO-sequencing data from Ebbing *et al.* (Ebbing *et al.* 2018) of worm slices from head to tail reveal specific expression in the head and tail regions for the neuron and pharynx cluster. (D-G) Analyzed TOMO-sequencing data (top panel) and RNA-sequencing data set on genes during development in *C. elegans* by Hendriks *et al.* (bottom panel) on the genes of each of the four groups: Cluster 1 (the neuron cluster genes), Cluster 2 (the pharynx cluster genes), genes that are expressed in the pharynx but not included in the pharynx cluster, and genes with the same oscillation phase as the average phase of the pharynx cluster.

The expression pattern enrichment of the pharynx cluster showed a significant enrichment for pharynx tissue, neuron and muscle-related tissue (Fig. 2B). The TOMO-sequencing data of sliced individual animals showed a strong enrichment of the pharynx cluster genes expression in the pharynx region (data shows suggestively two bulbs), while almost no expression was detected in other tissues (Fig. 2C-E). This finding is consistent with expression pattern enrichment, since the pharynx consists out of several types of tissue e.g. muscle and a few neurons. Phenotype and GO term enrichment analysis showed significant enrichment of eating variants and GO terms related to muscle activity, suggestive for the pumping activity of pharynx involving the pharyngeal muscles (Fig. 2B).

The pharynx cluster genes are specifically expressed in the pharynx region. Interestingly, our analysis showed that not all genes that are expressed in the pharynx are included in the pharynx cluster. Genes that are expressed in the pharynx region but that are not included in the pharynx cluster (retrieved from TOMO-sequencing data, see Methods), were more frequently expressed also outside of the pharynx region (Fig. 2F). Using a published RNA-sequencing data set on developmental gene expression dynamics in *C. elegans* (Hendriks et al. 2014), we found that the pharynx cluster consists solely out of genes which oscillate in a specific average phase of $253^\circ \pm 115^\circ$ (Fig. 2E). Pharynx genes that are not included in the pharynx cluster are also oscillating, and have a similar phase distribution with an average phase of $220^\circ \pm 126^\circ$ (Fig. 2F). Genes that have a similar phase in oscillations as the average phase of the pharynx cluster genes, are enriched for epithelial tissue (Fig. S2A) and show in the TOMO-sequencing data expression in the pharynx region but also in many other tissues (Fig. 2G). Finally, genes expressed in the pharynx but not included in the pharynx cluster or among genes with the same oscillation phase, do not show co-variance of expression among the samples with the pharynx cluster genes in our RNA-sequencing data (Fig. S2B). Taken together, the pharynx cluster represents a combination of genes that are both highly specifically expressed in the pharynx region and exhibit a specific phase of oscillations.

4.3.3 Spatial expression patterns and correlations between cluster genes

4.3.3.1 Analysis of the neuron cluster genes

To acquire more information on the spatial expression of the neuron cluster genes and to validate the existence neuron cluster during the young adult stage, we used the smFISH method (Raj et al. 2008) to correlate the expression between pairs of genes in the neuron cluster. The TOMO-sequencing data, as well as our own single animal RNA-sequencing data did not reveal in which cells or tissue the neuron cluster genes are expressed. The smFISH method enables us to get spatial information on the

expression of the cluster genes in individual cells and tissues, and also to quantify the mRNA levels for two genes simultaneously, allowing us to directly measure the same correlations in gene expression as for the RNA-sequencing analysis, but with high precision and single-cell resolution.

The neuron cluster is represented by three selected genes: *snt-1*, *far-8* and *daf-7*. *snt-1*, an ortholog of human SYT1 (synaptotagmin 1) and SYT2 (synaptotagmin 2), is involved in synaptic vesicle regulation (Nonet et al. 1998; Cheng et al. 2015), and is predicted to be mainly expressed in the head neurons. Indeed, we found that *snt-1* is broadly expressed in the head of the animal, i.e. neurons, but also in body wall muscles, pharyngeal cells and unknown cells (Fig. 3A). *far-8*, from the fatty acid- and retinol-binding protein (FAR) family has an unknown function in *C. elegans*. We showed that *far-8* is expressed in two amphid sheath cells surrounding the nerve ring (as identified based on expression of *vap-1*). Amphid sheath cells are known to be involved in maintaining the morphology of the receptive endings of a subset of sensory neurons (Bacaj et al. 2008). The well-described ligand *daf-7* from the TGF- β family is known to play a pivotal role in the behavioral regulation of *C. elegans*, including avoidance of pathogenic bacteria, feeding and foraging, as well as acquired transgenerational pathogenic avoidance behavior (Meisel et al. 2014; Milward et al. 2011; You et al. 2008; Moore, Kaletsky, and Murphy 2019). The *daf-7* cluster gene is known to be expressed in the OLQ, ASI, ASJ and ADE neurons (Meisel et al. 2014). These neurons can be distinguished from each other by their spatial expression pattern. We also found low levels of *daf-7* mRNA expression in cells with unknown identity close to the ASI and ASJ neurons.

For the *daf-7* gene, we could quantify expressions levels in all cells, with the option to measure each cell individually. In contrast, the expression levels of *far-8* were very high in the amphid sheath cells, causing overlap of the mRNA spots in the smFISH data, making it impossible to quantify individual mRNAs. Instead, we measured the average smFISH intensity levels of *far-8* in the amphid sheath cells. The cluster gene *snt-1* also has high expression levels. However, here mRNA molecules were distributed over the entire head region, causing less issues for individual mRNA detection. We quantified *snt-1* mRNA levels in a region that includes the whole nerve ring, terminal bulb, metacarpus and the surrounding tissue.

To test for correlations between the three candidate genes, we simultaneously quantified two cluster genes in individual animals. Only animals with a body length of at least 800 μ m (approximately after the fourth molt) are included, corresponding to the young adult stage used in our RNA-sequencing data. Expression of *snt-1* and *far-8* showed a significant but weak correlation ($p=0.05$) (Fig. 3B-C). For *far-8*, only the intensity within the amphid sheath cell area was measured, instead of absolute

mRNA levels, causing our correlations with *far-8* to be less reliable. Unfortunately, we do not have data with *daf-7* and *snt-1*, since the specific experimental batch returned poor data quality unrelated to the probe quality. We found that all three genes showed variable expression levels but no obvious expression dynamics with the body length. These observations indicate, in agreement with the RNA-sequencing analysis in Fig. 1C, that genes in the neuron cluster show little expression dynamics over the course of development (Fig 3D).

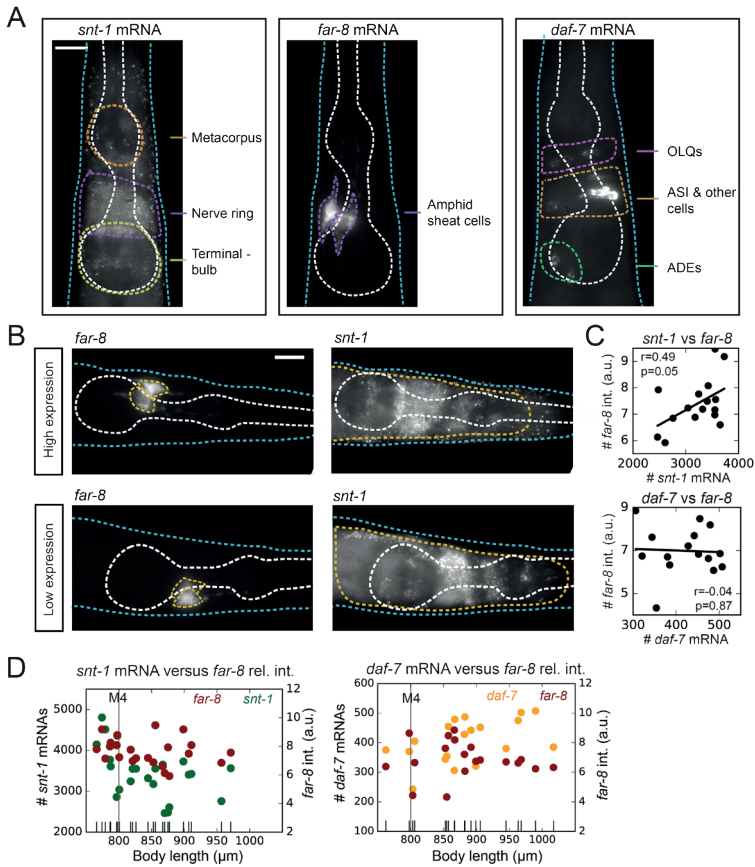


Figure 3: Neuron cluster genes show weak co-variance in smFISH data. (A) Microscopy images of the three neuron cluster genes, *snt-1*, *far-8* and *daf-7*. Light blue lines: body outline, white lines: pharynx outline. Other lines indicate different anatomical structures in the head. Scale bar 15mm. **(B)** Expression of the neuron cluster genes pair *snt-1* and *far-8* shows a significant, but weak correlation. By eye, only a slight difference is visible between animals that have a low *far-8* and *snt-1* expression compared to animals with high expression levels of *far-8* and *snt-1*. Light blue lines: body outline, white lines: pharynx outline, yellow lines: region of interest. Scale bar 15mm.

(C) Correlation plots of *snt-1* versus *far-8*, and *daf-7* versus *far-8*, where the first gene pair shows a weak significant correlation. Significance of correlations was measured with Spearman's rank correlation. (D) Expression levels of gene pairs as function of body length for individual animals, showing that expression levels do not depend strongly on developmental stage. Black insets on the x-axis represent the individual animals.

To summarize, we found a significant, but weak correlation in expression between the neuron cluster genes *snt-1* and *far-8*. However, it was difficult to quantify the expression levels of *far-8*, making the correlation analysis potentially less reliable. Unfortunately, we do not have smFISH data on *snt-1* and *daf-7* correlations. Overall, the correlations we found with smFISH for the neuron cluster are weaker than expected based on RNA-sequencing data.

4.3.3.2 Analysis of the pharynx cluster genes

All three pharynx cluster genes, *nhr-7*, *marg-1* and *tnt-4* are exclusively expressed in the pharynx region, as expected from the TOMO-sequencing data (Fig. 4A). The pharynx cluster genes *nhr-7* and *tnt-4* are both expressed in the muscle cells of the pharynx. The pharynx muscle-specific troponin *tnt-4* is a regulator of calcium in sarcomeres of the pharyngeal muscle required for muscle contraction. Removal of *tnt-4* with RNAi resulted in defects in the gland cells, which are embedded in the muscle/marginal cell epithelium and are thought to aid in digesting food (Raharjo et al. 2011). The *nhr-7* gene is nuclear hormone receptor and presumably a transcription factor. Nuclear hormone receptors affect a broad scale of process, such as, developmental timing, diapause and lifespan (Taubert, Ward, and Yamamoto 2011). However, there are no known phenotypes when *nhr-7* is removed. *nhr-7* is also expressed in the amphid sheath cells nearby the pharynx, as determined by *vap-1* expression (data not shown), similar to the expression of *far-8* from the neuron cluster. In addition, it is expressed in one unknown cell with a similar location and size as the amphid sheath cells but opposite side. The pharynx cluster gene *marg-1* is solely expressed in cells located in between the pharyngeal muscle cells, consistent with the position of the marginal cells of the pharynx. The marginal cells are involved in savaging important nutrients, e.g. cholesterol and other hydrophobic small molecules (Kamal et al. 2019).

Quantification of the whole pharynx region in animals of young adult age is challenging, because the pharynx is larger than the field of view of the microscope, requiring stitching of images and processing large regions for smFISH quantification. Instead, we quantified the expression levels in a specific region of the pharynx, namely the pm6 and pm5 cells in case of *nhr-7* and *tnt-4*, and the mc2 cells in the

case of *marg-1*. Indeed, for all three genes the mRNA levels quantified within this region appeared representative of mRNA levels found in the whole pharynx (Fig. S3A). Measured this way, we found a moderately significant ($p=0.03$) correlation between *nhr-7* and *marg-1* expression within the age range of the RNA-sequencing data (Fig. 4B-C). Higher mRNA expression levels of *marg-1* and *nhr-7* are mainly found right after the fourth molt, consistent with the expression peak around the fourth molt predicted from the oscillation observed by RNA-sequencing analysis (Fig. 4D, 2E). The expression levels of *tnt-4* were too high for distinguishing individual mRNAs. Instead, we measured the average intensity of the area of the pharynx as an indication of the gene expression levels. As a consequence, these measurements are less accurate than *nhr-7* and *marg-1* expression levels. Nevertheless, the resulting correlation between *nhr-7* and *tnt-4* is significant ($p=0.012$) (Fig. 4C).

In analyzing correlations in gene expression, we only included animals with body length larger than 800mm, corresponding to animals around the fourth molt. When animals that are younger are included, the correlation between the gene pairs becomes stronger (Fig. S3B). During the fourth molt, the expression levels of the three pharynx cluster genes are strongly decreased, as expected based on the predicted oscillation during the fourth molt (Fig. 2A). There is no evidence that animals in their fourth molt were included in our RNA-sequencing data, given that all animals included in the RNA-sequencing experiment had a measured body length of above ~780mm. Also, molting genes whose expression is expected to peak right before the fourth ecdysis had a relatively low number of reads in every animal in our data set, indicating that this molt had indeed already taken place (Fig. S3C, see Methods).

Taken together, we found a moderate significant correlation between the gene pairs *marg-1* and *nhr-7*, and *nhr-7* and *tnt-4* in staged young adults. All three pharynx cluster genes show increased expression levels right after the fourth molt. Here we show that the predicted oscillatory dynamics in time of the genes can be observed up until right after the fourth molt. Moreover, including younger animals increased the correlation between the gene pairs. However, there are no indications younger animals were included in the RNA-sequencing experiment. Overall, the pharynx cluster seems to be valid, even though the correlations measured by smFISH appeared weaker than those measured based on our RNA-sequencing data.

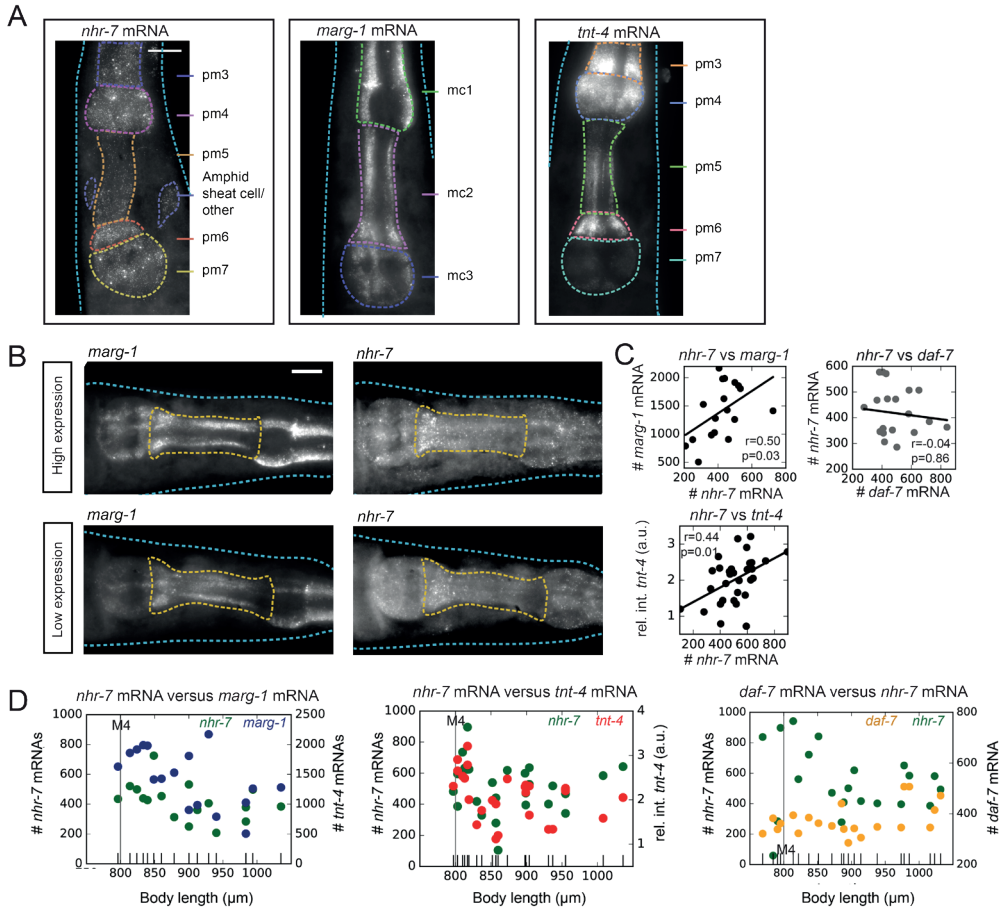


Figure 4: Pharynx cluster genes show slight to moderate co-variance in smFISH data. (A) Microscopy images of the three pharynx cluster genes, *nhr-7*, *marg-1* and *tnt-4*. Tissue and cells of which identity is known are annotated in colors, where pm cells are muscle specific cells in the pharynx and mc cells are marginal cells. Light blue lines: body outline. Scale bar 15mm. **(B)** Comparison of animals with low expression levels of *nhr-7* and *marg-1* to animals with high expression levels of *nhr-7* and *marg-1*. Light blue lines: body outline, yellow lines: region of interest. Scale bar 15mm. **(C)** Correlation plots of *nhr-7* versus *marg-1*, and *nhr-7* versus *tnt-4*, where we found significant correlations in expression both for the first gene pair ($p=0.03$) and the second ($p=0.012$). Significance of correlations was measured with Spearman's rank correlation. There is no significant correlation between the pharynx and neuron cluster genes, *nhr-7* and *daf-7*. **(D)** Expression levels of gene pairs as function of the body length of the individual animals. The expression dynamics is consistent with oscillatory gene expression, with a peak of expression right after the fourth molt (third panel, *nhr-7*).

4.3.4 Transgenerational memory effect on starvation cannot explain the correlations within or between the clusters

The correlations between gene pairs within the neuron and pharynx cluster or between both clusters, if significant, were less pronounced than measured for the RNA-sequencing data (Fig. 3C, 4C). Our analysis of RNA-sequencing data on transgenerational inheritance of starvation history (Webster et al. 2018), showed that the neuron cluster genes (and less pronounced the pharynx cluster genes) are upregulated in animals whose great-grandparents were subjected to prolonged severe starvation in the dauer state (Werner et al. 2020) (Fig. S4). This finding suggests that the expression levels of the cluster genes are subjected to starvation history of the animals, which could explain increased variance of gene expression between animals in case the inheritance of the transgenerational memory of starvation is stochastic among individual animals. Moreover, the neuron cluster gene *daf-7* has been shown before to be involved in transgenerational behavior memory to pathogenic bacteria (Moore, Kaletsky, and Murphy 2019). Our previous smFISH experiments on the clusters used animals under well-fed conditions for several generations, meaning that such transgenerational starvation memory effects play no role. However, the experimental protocol followed for the RNA-sequencing data did not rule out starvation in earlier generations, and we hypothesized that this might explain the difference in correlation observed between the smFISH and RNA-sequencing data. Therefore, we examined if animals of which previous generations have been subjected to severe starvation, resulted in more pronounced correlations when measured by smFISH.

To study the contribution of a transgenerational starvation memory effect on the co-variation of the clusters, we performed smFISH expression analysis on young adults whose ancestors were severely starved (Fig. 5A). We first starved animals (P0) for at least one month (as dauers on plates), after which they were put on food to recover. Subsequently the first generation (F1) was re-plated on food. Finally, the second generation (F2) was used for the smFISH experiments. From the data generated in the Webster *et al.* study, we expected an increase in expression levels of the cluster genes in the F2 group with starved ancestors compared to the control group (Fig. S4). However, we observed no significant increase in expression levels in the F2 group with starved ancestors compared to the control group for the cluster genes, *nhr-7*, *daf-7* and *tnt-4* (Fig. 5B). Also, the correlation between the two clusters represented by *nhr-7* (pharynx cluster) and *daf-7* (neuron cluster) showed no significant correlation, similar to the result shown previously in the well-fed control conditions (Fig. 5C). However, on food *daf-7* has significantly decreased expression levels in the F2 group compared to the control group. To find out what the likely identity of the cells are that have decreased expression levels during starvation,

we spatially grouped the cells that express *daf-7* into three categories, OLQs, ASI/other and ADE (Fig. 5E). Here, the ASI/other and OLQs in the F2 group on food have slightly decreased *daf-7* expression levels compared to control on food. Thus, even though, we observed no increase in expression of *daf-7*, we do observe that the expression levels are slightly modulated in animals with starved ancestors under control condition.

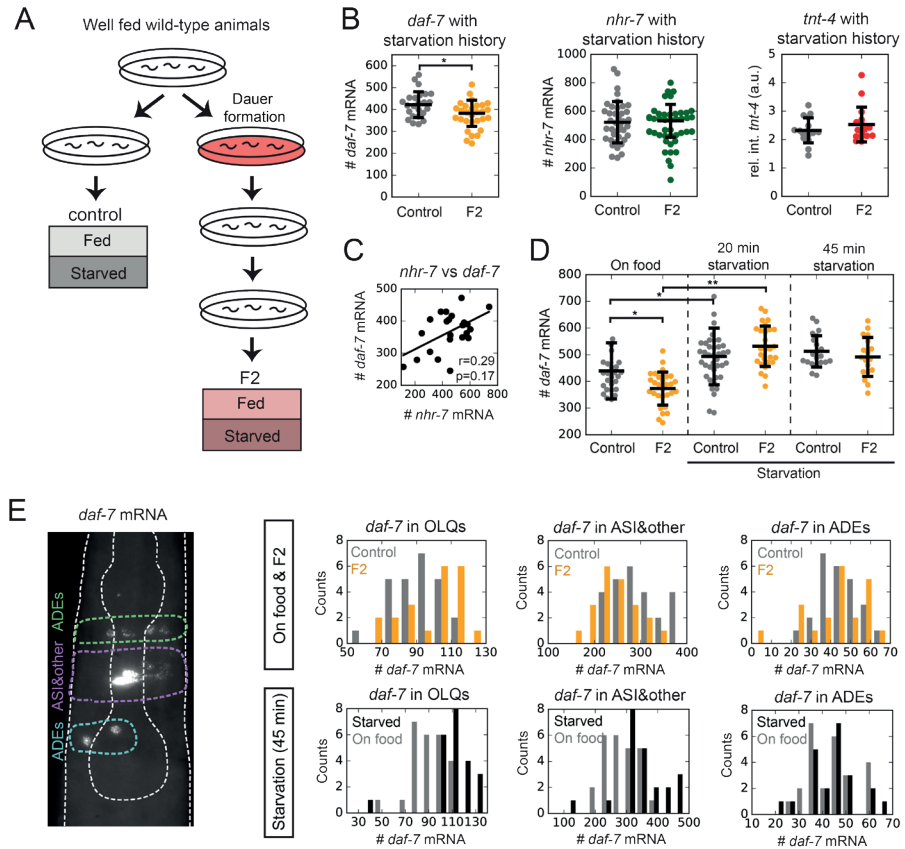


Figure 5: Transgenerational memory effect on starvation does not increase expression correlations within or between clusters. (A) Well-fed animals are either kept on food or starved for a prolonged time (>1 month) on plates as dauers. Dauers were replated on food and after two generations (F2) used for smFISH experiments to test transgenerational memory effects on starvation. (B) Comparison of cluster genes *daf-7*, *nhr-7* and *tnt-4* expression levels between control animals and F2 animals with starved ancestors on food. The cluster gene *daf-7* shows a significant transgenerational memory effect due to starvation that leads to lower expression levels. (C) Expression levels of the two cluster genes *nhr-7* and *daf-7*, representing the neuron and pharynx cluster, respectively, measured in individual animals. The correlation in expression between the

two genes is not significant. **(D)** Starvation induces an increase in *daf-7* expression in both the F2 and control group after 20 and 45 minutes of starvation. **(E)** The expression levels of *daf-7* in OLQ, ASI/other and ADE neurons, in animals on food and starved, and with without ancestral starvation history. The significant decrease in expression levels in animals with starved ancestors compared to the control group, is caused mainly by the decreased expression levels in OLQ and ASI (/other cells) neurons. When animals are starved for 45 minutes, the OLQs and ASI/other cells have increased expression compared to control. Significance of expression difference assessed by a pair-wise t-test, * $P < 0.05$, and ** $P < 0.01$.

During the RNA-sequencing experiments, the animals did not have access to food for ~45-60 minutes before sequencing. *daf-7* is known to be involved in entering the dauer stage (Swanson and Riddle 1981; Vowels and Thomas 1992), and *daf-7* has been described as key player in behavioral feeding response (Hilbert and Kim 2017). When well-fed animals are starved for 45 minutes, the OLQs, ASI/other show increased *daf-7* expression levels compared to animals on food (Fig. 5E). Since *daf-7* expression levels are slightly modulated in animals with starved ancestors under control condition, we wondered if *daf-7* reacts differently on starvation when animals have a different feeding history. We find an increase in expression of *daf-7* when the animals are starved already after 20 minutes in the control group (Fig. 5D). However, we did not find a difference in the response of *daf-7* to starvation after 20 or 45 minutes in absence of food between control and animals with starved ancestors. Hence, starvation history of ancestors does not have an effect on the starvation response for *daf-7* or *nhr-7* in animals two generation later.

To summarize, we did not find increased expression levels of *daf-7*, *nhr-7* or *tnt-4*, when we subjected animals to a transgenerational starvation memory effect. Also, the correlation within or between the cluster did not change. However, we did find a transgenerational memory effect on *daf-7* expression levels in the F2 generation when on food. This shows that there is a transgenerational memory effect in at least one gene from the neuron cluster. But this finding appears not directly related to the presence of the clusters in the RNA-sequencing data.

4.3.5 Direct starvation response induces correlation between clusters due to age related upregulation of gene expression

Transgenerational starvation memory effects did not induce a stronger correlation between or within the clusters. In the starvation study by Harvald *et al.* (Harvald *et al.* 2017), animals were also starved for several hours from mid-L4 stage onward leading to identification by RNA-sequencing of a large set of genes upregulated under direct starvation. Analysis of this data set revealed that almost all genes in the pharynx and neuron cluster are predicted to increase expression upon starvation of 1-16

hours (Fig. S5). Animals in our RNA-sequencing experiments were starved between 45-60 minutes before they were fixed for sequencing. If the starvation response would differ between the animals, e.g. due to variations in developmental age or a stochastic starvation response, then this could explain the correlated variability in gene expression observed within and between the clusters we identified by RNA-sequencing.

To address the direct starvation response in the cluster genes, we starved the animals for 45 minutes and fixed them immediately for smFISH experiments. The genes *nhr-7* from the pharynx cluster and *daf-7* from the neuron cluster allowed most reliable quantification in smFISH experiments, due to their relatively low expression levels. Indeed, *daf-7* and *nhr-7* showed increased expression levels when starved for 45 minutes compared to the well-fed control group (Fig. 6A). An increase in expression levels upon 45 minutes of starvation was not observed in *marg-1*, suggesting that *marg-1* reacts differently to starvation than *nhr-7* from the pharynx cluster. Surprisingly, increased expression levels of *daf-7* upon starvation only occurred in animals within a specific body length-range (Fig. 6A, 4D), corresponding to the stage of development directly after the fourth molt. The increase in expression of *daf-7* upon starvation happens at a specific developmental age, which at the same developmental as when we see an increase in expression of the pharynx cluster genes due to oscillating gene expression. The correlated variation in the RNA-sequencing data, that could not be reproduced by our previous results, can be explained by the age dependent increase of expression upon the starvation of the neuron cluster at the exact same developmental time of a peak of expression of the genes from the pharynx cluster.

Taken together, the expression levels of *daf-7* and *nhr-7* are increased when starved for 45 minutes. The 45 minutes starvation response of *daf-7* is highest at a specific developmental stage, right after the fourth molt. At the same developmental stage, the *nhr-7* and *marg-1* also have increased expression levels due to gene expression oscillations related to development. This increase of expression levels at a specific developmental stage, gives a higher correlation between both clusters compared to previous correlations found in well-fed animals or in animals with starvation history. In summary, the induced starvation response in the samples could explain the correlation found between both cluster in the RNA-sequencing data.

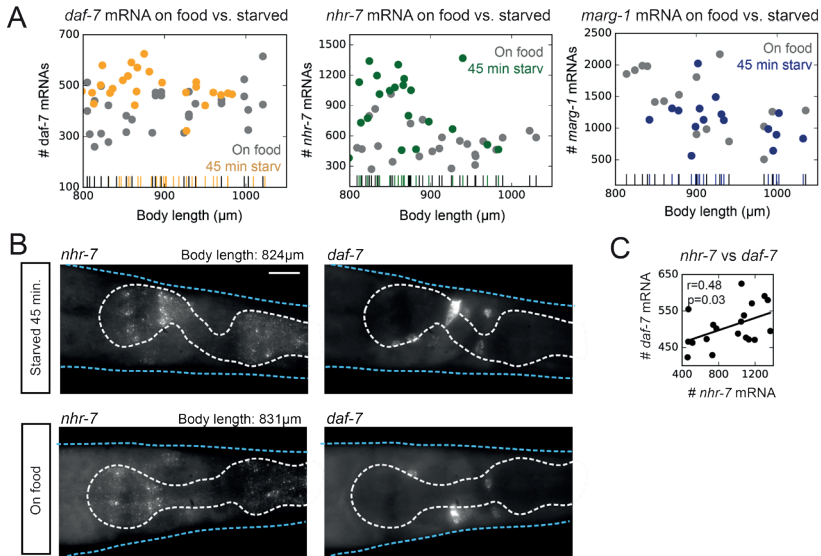


Figure 6: Starvation response induces correlation between the two cluster genes *daf-7* and *nhr-7*. (A) Expression levels of cluster genes *daf-7*, *nhr-7* and *marg-1* as function of body length in individual animals, divided into two groups: on food versus starved for 45 minutes. The data on *daf-7* is the same as in Fig. 5D. (B) Animals with similar body length on food and starved for 45 minutes. Starved animals show an increased expression of *daf-7* and *nhr-7* at the developmental stage right after fourth molt. Light blue lines: body outline, white lines: pharynx. Scale bar 15 μm. (C) Significant correlation between *daf-7* and *nhr-7*, representative for the neuron and pharynx cluster respectively, when animals are starved for 45 minutes.

4.4 Discussion

The aim of this study was to provide insight into the nature and function of two clusters of genes, the pharynx and neuron cluster, that were found with the novel density-based clustering approach on single animal RNA-sequencing data of young adult *C. elegans* hermaphrodites (Werner et al. 2020). Both clusters consist out of genes that have correlated expression between each other and between the cluster themselves. We selected for each cluster three representative genes of which expression levels were quantified with smFISH (Raj et al. 2008) to test for correlations in gene expression for gene pairs within and between the clusters. The pharynx cluster is represented by *nhr-7*, *marg-1* and *tnt-4*. We found significant correlations between gene pairs for the pharynx cluster, meaning that the pharynx cluster in the RNA-sequencing data is valid. Moreover, the genes were predicted to oscillate during the late L3 and L4 stage (Hendriks et al. 2014). We showed that the expression of the selected genes indeed oscillates, where the peak related to the oscillations is located

right after the fourth molt, within the age range of the RNA-sequencing animals. Our findings identified the oscillations in expression levels observed in the genes of the pharynx cluster as the most likely cause of the correlated variability in expression of the genes in the pharynx cluster we observed in our RNA-sequencing analysis.

Next, we checked for correlated expression of three selected genes from the neuron cluster, *far-8*, *snt-1* and *daf-7*. We found a significant, but weak correlation between *far-8* and *daf-7* expression, and no significant correlation between *far-8* and *snt-1*. Unfortunately, the expression of *far-8* was hard to quantify with the smFISH method, causing less reliability in the measurements. Also, in the process of these measurements, *snt-1* was excluded from the neuron cluster after we applied stricter rules in the clustering approach (Werner et al. 2020), resulting in a degree of uncertainty whether *snt-1* is representative of the neuron cluster. Last, *daf-7* included several zero read counts in the samples of the RNA-sequencing data (12/34), as an indication that the measurement could be noisy. Taken together, our results on correlated expression of gene pairs in the neuron cluster are not reliable enough to make a conclusive statement about the validation of the neuron cluster.

Analysis on the RNA-sequencing data on transgenerational inheritance of starvation history of the study from Webster *et al.* (Webster et al. 2018), revealed that the expression levels of neuron and of several pharynx cluster genes are increased under well-fed conditions when animals have an ancestral starvation history (third generation). With smFISH, we showed that the expression levels of *daf-7*, *nhr-7* and *tnt-4* are not increased upon ancestral starvation history under well-fed conditions. Moreover, closer inspection of the RNA-sequencing data from Webster *et al.* on the third-generation animals with an ancestral starvation history, revealed unexplained biological related differences or measurements inconsistencies between the samples (data not shown). The reliability of the data set from Webster *et al.*, could be influenced by this inconsistency. Nevertheless, we did show a difference between well-fed control and ancestral starvation history in *daf-7* expression levels. Animals in which great-grand parents were severely starved, *daf-7* expression levels were significantly decreased and the expression was modulated mainly in the ASI neurons (and closely surrounding cells) and weakly in the OLQ neurons. The ligand *daf-7* from the TGF- β family, is known to be involved in acquired transgenerational pathogenic avoidance behavior (Moore, Kaletsky, and Murphy 2019). Moreover, *daf-7* is also involved in regulation of other behaviors, such as feeding, foraging, quiescence, and aggregation tendency (Ben Arous, Laffont, and Chatenay 2009; Chang et al. 2006; Greer et al. 2008), suggesting a possible change in behavior in animals with ancestral starvation history. It would be interesting to find out if the behavior of the animals with ancestral starvation history is different under specific conditions.

The animals in our RNA-sequencing data could potentially have a starvation response, since the animals have been in absence of food for 45-60 minutes before they were fixed for sequencing. We showed that under starvation conditions of 45 minutes, *daf-7* expression levels increase in the ASI (and surrounding cells) and OLQ neurons. Moreover, the increase of *daf-7* in the ASI neurons has been shown before in animals under 1 day of dietary restriction (Fletcher and Kim 2017). The expression levels of *nhr-7* also increased upon starvation. In contrast, expression levels of *marg-1* did not change upon starvation. *nhr-7*, *marg-1* and *daf-7*, were all expected to increase expression levels, based on the analysis of the RNA-sequencing data from the starvation study by Harvald *et al.* (Harvald et al. 2017). The developmental age of the animals could explain this discrepancy, since in the starvation study the animals were starved at mid-L4 stage and the pharynx genes are oscillating until right after the fourth molt, suggesting that the observed increase in expression levels could be related to the oscillations.

In the starvation experiments with *daf-7*, we observed an interesting dependence on developmental stage of expression: animals with a developmental stage right after the fourth molt showed increased *daf-7* expression levels. The increased expression levels of *daf-7* specific to early-L4 animals occurs at the same developmental stage as the peak of expression in the oscillating genes from the pharynx cluster. The correlated variation in the RNA-sequencing data, could be explained the age dependent increase of expression upon starvation of the neuron cluster, at the same developmental age of the peak of expression in the pharynx cluster. The developmental age at which the *daf-7* expression is increased could be related to the checkpoint at which animals arrest their development (Schindler, Baugh, and Sherwood 2014). Since, part of the neuron cluster already partly consists out of genes that show oscillatory-like behavior with the same phase as the pharynx cluster (12/56), the genes without these expression dynamics could potentially show similar oscillatory-like behavior right after the fourth molt upon starvation.

The pharynx cluster consists out of genes with highly specific spatial expression and a specific phase of oscillations. It remains unclear why other pharynx genes with similar oscillatory phase dynamics were not included into the pharynx cluster. Pharynx genes that were not included into the pharynx cluster, show stronger variation and no correlation with the pharynx cluster. The exclusion of the other pharynx genes could be related to the confined area of the pharynx, in which genes have more similar expression levels and oscillations phases than genes expressed broadly throughout the body. Moreover, oscillating genes with broader expression could have oscillations outside the pharynx with a different behavior at the young adult stage than genes in the pharynx region. Also, the reaction to starvation could potentially cause differences in starvation response among different tissues.

While our work validated our initial identification of the pharynx cluster, it did not validate the neuron cluster. However, this was at least partly caused by shortcomings in the experimental design, since two out of the three selected genes from the neuron cluster either lacked strong evidence for belonging to the neuron cluster or could not be quantified with smFISH. Nevertheless, the novel clustering approach found very specific groups of genes, with clear common features shared by the genes within each cluster. The correlation found between both clusters in the RNA-sequencing data could be explained by the induced starvation response. Moreover, *daf-7* from the neuron cluster increased expression levels upon starvation at the same developmental stage as when the expression peaks of the pharynx cluster genes due to the oscillatory behavior. However, the biological relevance of the dependence on developmental stage of the response of *daf-7* expression upon starvation, as well as the discovered modulation of its expression in animals with ancestral starvation history, remain to be explored further. For example, behavioral assays on animals with ancestral starvation history could provide more insight into the potential behavioral changes.

4.5 Material and methods

C. elegans strains and handling

Wild-type nematodes were strain N2. No mutants were used in this study.

All strains were handled according to the standard protocol (Brenner 1974). Briefly, animals were grown on agar plates, consisting out of nematode growth medium (NGM), seeded with the *E. coli* strain OP50 as food source and grown at 20 °C.

Dauer induction

To obtain animals in dauer stage (in order to obtain animals with an ancestral starvation history), between 8-10 N2 animals were transferred to a freshly seeded NGM plate with OP50. Animals were grown and subsequently starved when animals run out of food on the plates for ~1.5 months. The dauer staged animals were retrieved by chunking a small part of the plate and transferring the chunk to a new freshly seeded plate. After ~1 hour, larger animals showing bagged fate, crawling on the plate were removed, to decrease the number of animals with of other stages than dauer.

Starvation induction

Starvation was induced by washing the animals from plates with M9 buffer and transferred into a 15 mL tube. After that, the animals were washed 3x with M9 buffer

for 5 minutes. The timer of starvation was started at the first wash step. After washing the bacteria of the animals, the animals were kept in the tube with M9 buffer, gently rocking until the starvation time was reached.

Single molecule fluorescence in situ hybridization

The oligonucleotides for the smFISH probe sets were designed with optimal GC content and specificity for the gene of interest using the Stellaris RNA FISH probe designer. The oligonucleotides were synthesized with a 3' amino C7 modification and purified by LGC Biosearch Technologies. Conjugation of the oligonucleotides with either Cy5 (GE Amersham) or Alexa594 (Invitrogen) was done as previously described (Lyubimova et al. 2013). The smFISH protocol was performed as previously described (Raj et al. 2008; Ji and van Oudenaarden 2012). Briefly, staged animals were washed from plates with M9 buffer and fixed in 4% formaldehyde in 1x PBS, gently rocking at room temperature (RT) for 40 minutes (young adults for 35 minutes). Fixation of embryos required a snap-freeze step to crack the eggshells by submerging embryos, after 15 minutes in fixation buffer, in liquid nitrogen, and thawing on ice for 20 minutes. After fixation, the animals were 2x washed with 1xPBS and resuspended in 70% ethanol overnight at 4°C. Ethanol was removed and animals were washed with 10% formamide and 2X SSC, as preparation for the hybridization. Animals were incubated with the smFISH probes overnight in the dark at 37°C in a hybridization solution (Stellaris) with added 10% formamide. The next day, animals were washed 2x with 10% formamide and 2X SSC each with an incubation step of 30 minutes at 37°C. The last wash step contains DAPI 5 µg/mL for nuclear staining. The wash buffer was removed, and animals were resuspended in 2X SSC and stored at 4°C until imaging. The 2X SSC was aspirated and animals were immersed in 100 µl GLOX buffer (0.4% glucose, 10 mM Tris-HCl, pH 8.0, 2X SSC) together with 1 µl Catalase (Sigma-Aldrich) and 1 µl glucose oxidase (Sigma-Aldrich) (3.7 mg/mL) to prevent bleaching during imaging.

Microscopy images of smFISH samples were acquired with a Nikon Ti-E inverted fluorescence microscope, equipped with a 100X plan-apochromat oil-immersion objective and an Andor Ikon-M CCD camera controlled by µManager software (Edelstein et al. 2014). smFISH analysis was performed with custom Python software, based on a previously described method (Raj et al. 2008). Briefly, we first convolved the smFISH images with a Gaussian filter. Next, candidate spots were selected via manual thresholding, and partially overlapping spots were separated via finding 3D regional intensity maxima.

Data analysis and handling

Published data sets

Single-animal RNA-sequencing data of Werner *et al.* (Werner et al. 2020), was analyzed and handled similarly as described in the study, with Python and R.

The RNA-sequencing data of Webster *et al.* (Webster et al. 2018), on transgenerational inheritance of starvation history was analyzed similarly as described in the study. We used the *edgeR* package (3.30.3) in R to normalize the data.

GO term enrichment

The tissue and phenotype enrichment were performed with the Enrichment analysis tool on Wormbase (Angeles-Albores et al. 2018; Angeles-Albores et al. 2016), and the GO term enrichment was conducted with the Panther Tool (Mi et al. 2019). In both cases, we entered the gene set to find annotated terms that were over-represented, compared to the default available background list on *C. elegans*.

4.6 Appendix

4.6.1 Author contribution to the chapter

JJHT, SW, JSvZ and TSS designed the experimental strategy. SW, JJHT, JSvZ and TSS conceptualized the study. JJHT and DvB performed the experiments. JJHT, DvB analyzed the smFISH data. JJHT and SW analyzed publicly available data sets. JJHT wrote the text. JSvZ reviewed and edited the text. JSvZ, SW and TSS supervised the study.

4.6.2 Supplementary figures

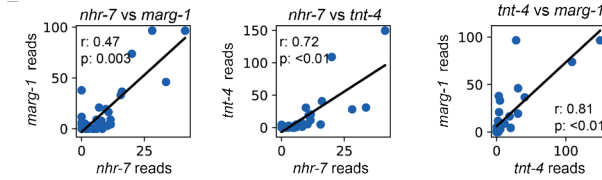


Figure S1: RNA-sequencing data without sample that appeared as an outlier. Correlations between the selected pharynx cluster genes in the RNA-sequencing data, in which we left out the sample which appeared as an outlier in the data. Without the sample, the selected pharynx cluster gene pairs still show significant ($p < 0.01$) correlations. Significance of correlations was measured with Spearman's rank correlation.

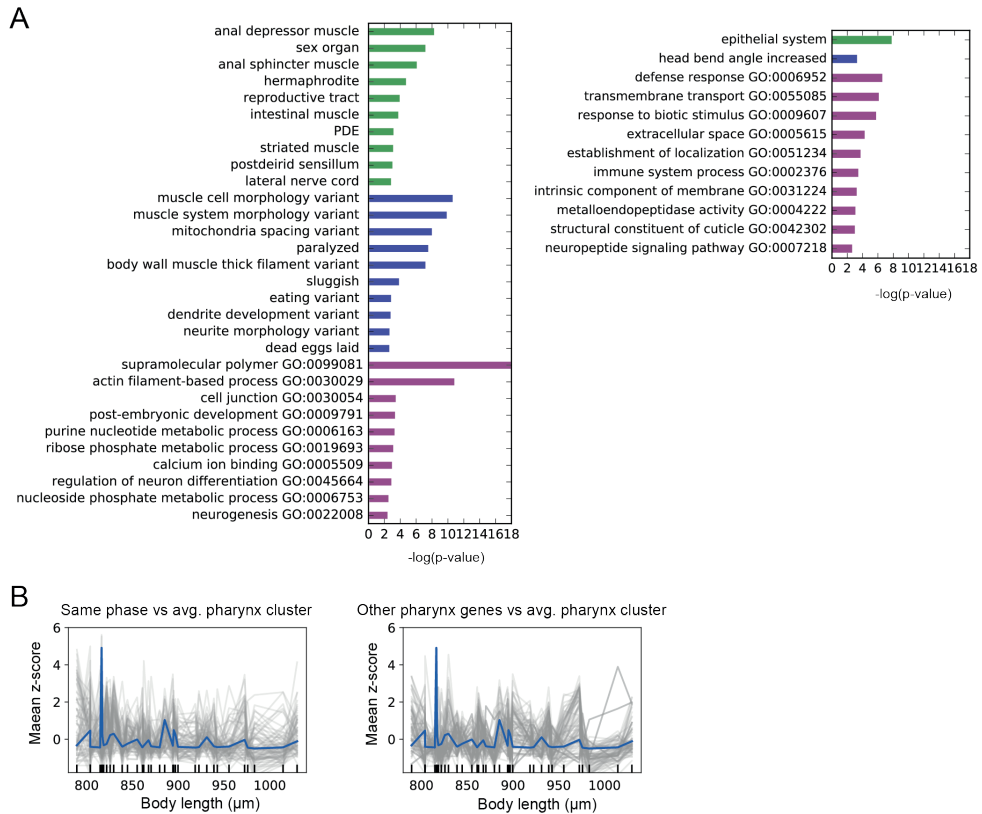


Figure S2: Enrichment analysis on genes with a similar phase as the pharynx cluster and other pharynx genes. (A) Tissue (green), phenotype (blue) and GO-term enrichment (purple) on other genes that are also expressed in the pharynx (left panel), and on genes with the same phase as the pharynx cluster (right panel). (B) Co-variance of the two groups: a group of genes with a similar phase as the pharynx cluster, and a group consisting out of genes that are also expressed in the pharynx but are not in the pharynx cluster. Both groups of genes do not show co-variance of expression among the samples with average mean z-score over all genes in the pharynx cluster (Blue line). Black insets on the x-axis represent the individual animals.

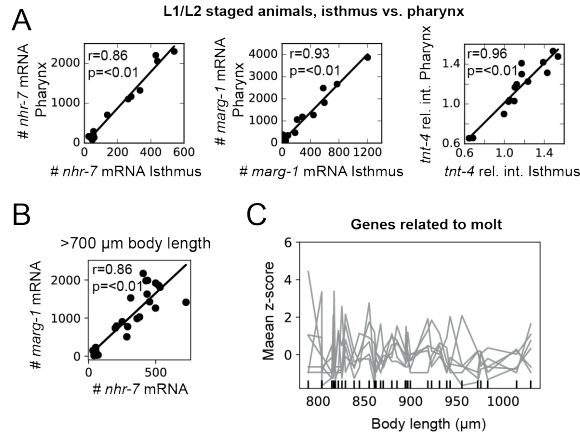


Figure S3: Increased correlation when younger animals are included than the age range of animals in the RNA-sequencing data. (A) Highly significant correlations between the selected region of the isthmus and the whole pharynx for the panel of genes from the pharynx cluster, *nhr-7*, *marg-1* and *tnt-4*, in L1/L2 staged animals. (B) Stronger correlation between *nhr-7* and *marg-1* when younger animals are included that are in the fourth molt, when the gene expression of all three genes is expected to be lower due to the oscillatory dynamics of the gene expression. Significance of correlations was measured with Spearman's rank correlation. (C) None of the genes related to molting show no dynamics in expression between the samples of the RNA-sequencing data, suggesting that the animals are already past the molting cycle. Black insets on the x-axis represent the individual animals.

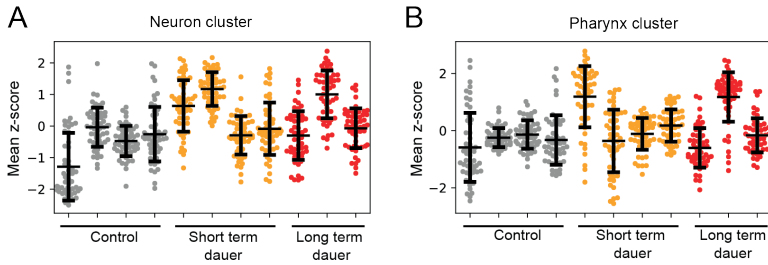


Figure S4: Analysis of RNA-sequencing data on transgenerational inheritance of starvation history on the pharynx and neuron cluster. (A-B) Our analysis on the RNA-sequencing data of Webster *et al.* (Webster *et al.* 2018), on transgenerational inheritance of starvation history for each of the clusters showed that the neuron cluster genes (A) and less pronounced the pharynx cluster genes (B), are upregulated in animals whose great-grandparents were subjected to prolonged severe starvation in the dauer state. The F3 generations were divided into two groups, dependent on their treatment: short term dauers (10 days) and long term dauers (36–45 days). Both groups show variability in gene expression between the replicates of each group.

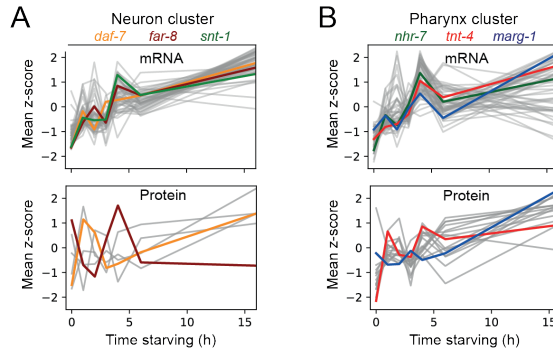


Figure S5: Analysis on starvation response of the neuron and pharynx cluster. (A-B) We analyzed the RNA sequencing data and proteomics data in the starvation study by Harvald *et al.* (Harvald et al. 2017), on animals that were starved from mid-L4 stage onward. Almost all gene in the neuron (A) and pharynx (B) cluster are predicted to increase expression upon starvation of 1-16 hours. The panel of genes for each cluster is depicted with a colored line.

4.7 References

- Anders, S., D. J. McCarthy, Y. Chen, M. Okoniewski, G. K. Smyth, W. Huber, and M. D. Robinson. 2013. 'Count-based differential expression analysis of RNA sequencing data using R and Bioconductor', *Nat Protoc*, 8: 1765-86.
- Angeles-Albores, D., R. Lee, J. Chan, and P. Sternberg. 2018. 'Two new functions in the WormBase Enrichment Suite', *MicroPubl Biol*, 2018.
- Angeles-Albores, D., N. Lee RY, J. Chan, and P. W. Sternberg. 2016. 'Tissue enrichment analysis for *C. elegans* genomics', *BMC Bioinformatics*, 17: 366.
- Bacaj, T., M. Tevlin, Y. Lu, and S. Shaham. 2008. 'Glia are essential for sensory organ function in *C. elegans*', *Science*, 322: 744-7.
- Ben Arous, J., S. Laffont, and D. Chatenay. 2009. 'Molecular and sensory basis of a food related two-state behavior in *C. elegans*', *PLoS One*, 4: e7584.
- Brenner, S. 1974. 'The genetics of *Caenorhabditis elegans*', *Genetics*, 77: 71-94.
- Chang, A. J., N. Chronis, D. S. Karow, M. A. Marletta, and C. I. Bargmann. 2006. 'A distributed chemosensory circuit for oxygen preference in *C. elegans*', *PLoS Biol*, 4: e274.
- Cheng, Y., J. Wang, Y. Wang, and M. Ding. 2015. 'Synaptotagmin 1 directs repetitive release by coupling vesicle exocytosis to the Rab3 cycle', *Elife*, 4.
- Dall, Jesper, and Michael Christensen. 2002. 'Random geometric graphs', *Physical Review E*, 66: 016121.
- Ebbing, A., A. Vertesy, M. C. Betist, B. Spanjaard, J. P. Junker, E. Berezikov, A. van Oudenaarden, and H. C. Korswagen. 2018. 'Spatial Transcriptomics of *C. elegans* Males and Hermaphrodites Identifies Sex-Specific Differences in Gene Expression Patterns', *Dev Cell*, 47: 801-13 e6.
- Edelstein, A. D., M. A. Tsuchida, N. Amodaj, H. Pinkard, R. D. Vale, and N. Stuurman. 2014. 'Advanced methods of microscope control using muManager software', *J Biol Methods*, 1.
- Fletcher, M., and D. H. Kim. 2017. 'Age-Dependent Neuroendocrine Signaling from Sensory Neurons Modulates the Effect of Dietary Restriction on Longevity of *Caenorhabditis elegans*', *PLoS Genet*, 13: e1006544.
- Gartner, K. 1990. 'A third component causing random variability beside environment and genotype. A reason for the limited success of a 30 year long effort to standardize laboratory animals?', *Lab Anim*, 24: 71-7.
- Greer, E. R., C. L. Perez, M. R. Van Gilst, B. H. Lee, and K. Ashrafi. 2008. 'Neural and molecular dissection of a *C. elegans* sensory circuit that regulates fat and feeding', *Cell Metab*, 8: 118-31.
- Grun, D., and A. van Oudenaarden. 2015. 'Design and Analysis of Single-Cell Sequencing Experiments', *Cell*, 163: 799-810.
- Han, Y., S. Gao, K. Muegge, W. Zhang, and B. Zhou. 2015. 'Advanced Applications of RNA Sequencing and Challenges', *Bioinform Biol Insights*, 9: 29-46.
- Harvald, E. B., R. R. Sprenger, K. B. Dall, C. S. Ejsing, R. Nielsen, S. Mandrup, A. B. Murillo, M. Larance, A. Gartner, A. I. Lamond, and N. J. Faergeman. 2017. 'Multi-omics Analyses of Starvation Responses Reveal a Central Role for Lipoprotein Metabolism in Acute Starvation Survival in *C. elegans*', *Cell Syst*, 5: 38-52 e4.
- Hendriks, G. J., D. Gaidatzis, F. Aeschimann, and H. Grosshans. 2014. 'Extensive oscillatory gene expression during *C. elegans* larval development', *Mol Cell*, 53: 380-92.
- Hilbert, Z. A., and D. H. Kim. 2017. 'Sexually dimorphic control of gene expression in sensory neurons regulates decision-making behavior in *C. elegans*', *Elife*, 6.

- Ji, N., and A. van Oudenaarden. 2012. 'Single molecule fluorescent in situ hybridization (smFISH) of *C. elegans* worms and embryos,' *WormBook*: 1-16.
- Kaern, M., T. C. Elston, W. J. Blake, and J. J. Collins. 2005. 'Stochasticity in gene expression: from theories to phenotypes,' *Nat Rev Genet*, 6: 451-64.
- Kamal, M., H. Moshiri, L. Magomedova, D. Han, K. C. Q. Nguyen, M. Yeo, J. Knox, R. Bagg, A. M. Won, K. Szlapa, C. M. Yip, C. L. Cummins, D. H. Hall, and P. J. Roy. 2019. 'The marginal cells of the *Caenorhabditis elegans* pharynx scavenge cholesterol and other hydrophobic small molecules,' *Nat Commun*, 10: 3938.
- Kirkwood, T. B., M. Feder, C. E. Finch, C. Franceschi, A. Globerson, C. P. Klingenberg, K. LaMarco, S. Omholt, and R. G. Westendorp. 2005. 'What accounts for the wide variation in life span of genetically identical organisms reared in a constant environment?,' *Mech Ageing Dev*, 126: 439-43.
- Luecken, M. D., and F. J. Theis. 2019. 'Current best practices in single-cell RNA-seq analysis: a tutorial,' *Mol Syst Biol*, 15: e8746.
- Lyubimova, A., S. Itzkovitz, J. P. Junker, Z. P. Fan, X. Wu, and A. van Oudenaarden. 2013. 'Single-molecule mRNA detection and counting in mammalian tissue,' *Nat Protoc*, 8: 1743-58.
- Meisel, J. D., O. Panda, P. Mahanti, F. C. Schroeder, and D. H. Kim. 2014. 'Chemosensation of bacterial secondary metabolites modulates neuroendocrine signaling and behavior of *C. elegans*,' *Cell*, 159: 267-80.
- Mi, H., A. Muruganujan, D. Ebert, X. Huang, and P. D. Thomas. 2019. 'PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools,' *Nucleic Acids Res*, 47: D419-D26.
- Milward, K., K. E. Busch, R. J. Murphy, M. de Bono, and B. Olofsson. 2011. 'Neuronal and molecular substrates for optimal foraging in *Caenorhabditis elegans*,' *Proc Natl Acad Sci U S A*, 108: 20672-7.
- Moore, R. S., R. Kaletsky, and C. T. Murphy. 2019. 'Piwi/PRG-1 Argonaute and TGF-beta Mediate Transgenerational Learned Pathogenic Avoidance,' *Cell*, 177: 1827-41 e12.
- Munsky, B., G. Neuert, and A. van Oudenaarden. 2012. 'Using gene expression noise to understand gene regulation,' *Science*, 336: 183-7.
- Newman, Mark. 2018. *Networks* (Oxford university press).
- Nonet, M. L., O. Saifee, H. Zhao, J. B. Rand, and L. Wei. 1998. 'Synaptic transmission deficits in *Caenorhabditis elegans* synaptobrevin mutants,' *J Neurosci*, 18: 70-80.
- Penrose, D.M.S.M., M. Penrose, and Oxford University Press. 2003. *Random Geometric Graphs* (Oxford University Press).
- Penrose, Mathew D., and Agoston Pisztor. 1996. 'Large Deviations for Discrete and Continuous Percolation,' *Advances in Applied Probability*, 28: 29-52.
- Perez, M. F., M. Francesconi, C. Hidalgo-Carcedo, and B. Lehner. 2017. 'Maternal age generates phenotypic variation in *Caenorhabditis elegans*,' *Nature*, 552: 106-09.
- Raharjo, W. H., V. Ghai, A. Dineen, M. Bastiani, and J. Gaudet. 2011. 'Cell architecture: surrounding muscle cells shape gland cell morphology in the *Caenorhabditis elegans* pharynx,' *Genetics*, 189: 885-97.
- Raj, A., P. van den Bogaard, S. A. Rifkin, A. van Oudenaarden, and S. Tyagi. 2008. 'Imaging individual mRNA molecules using multiple singly labeled probes,' *Nat Methods*, 5: 877-9.
- Raj, A., and A. van Oudenaarden. 2008. 'Nature, nurture, or chance: stochastic gene expression and its consequences,' *Cell*, 135: 216-26.
- Rando, O. J. 2012. 'Daddy issues: paternal effects on phenotype,' *Cell*, 151: 702-08.
- Saint, M., F. Bertaux, W. Tang, X. M. Sun, L. Game, A. Koferle, J. Bahler, V. Shahrezaei, and S. Marguerat. 2019. 'Single-cell imaging and RNA sequencing reveal patterns

- of gene expression heterogeneity during fission yeast growth and adaptation', *Nat Microbiol*, 4: 480-91.
- Schindler, A. J., L. R. Baugh, and D. R. Sherwood. 2014. 'Identification of late larval stage developmental checkpoints in *Caenorhabditis elegans* regulated by insulin/IGF and steroid hormone signaling pathways', *PLoS Genet*, 10: e1004426.
- Swanson, M. M., and D. L. Riddle. 1981. 'Critical periods in the development of the *Caenorhabditis elegans* dauer larva', *Dev Biol*, 84: 27-40.
- Taubert, S., J. D. Ward, and K. R. Yamamoto. 2011. 'Nuclear hormone receptors in nematodes: evolution and function', *Mol Cell Endocrinol*, 334: 49-55.
- Vowels, J. J., and J. H. Thomas. 1992. 'Genetic analysis of chemosensory control of dauer formation in *Caenorhabditis elegans*', *Genetics*, 130: 105-23.
- Wagner, A., A. Regev, and N. Yosef. 2016. 'Revealing the vectors of cellular identity with single-cell genomics', *Nat Biotechnol*, 34: 1145-60.
- Webster, A. K., J. M. Jordan, J. D. Hibshman, R. Chitrakar, and L. R. Baugh. 2018. 'Transgenerational Effects of Extended Dauer Diapause on Starvation Survival and Gene Expression Plasticity in *Caenorhabditis elegans*', *Genetics*, 210: 263-74.
- Werner, Steffen, W Mathijs Rozemuller, Annabel Ebbing, Anna Alemany, Joleen Traets, Jeroen S. van Zon, Alexander van Oudenaarden, Hendrik C. Korswagen, Greg J. Stephens, and Thomas S. Shimizu. 2020. 'Functional modules from variable genes: Leveraging percolation to analyze noisy, high-dimensional data', *bioRxiv*: 2020.06.10.143743.
- Wong, A. H., Gottesman, II, and A. Petronis. 2005. 'Phenotypic differences in genetically identical organisms: the epigenetic perspective', *Hum Mol Genet*, 14 Spec No 1: R11-8.
- You, Y. J., J. Kim, D. M. Raizen, and L. Avery. 2008. 'Insulin, cGMP, and TGF-beta signals regulate food intake and quiescence in *C. elegans*: a model for satiety', *Cell Metab*, 7: 249-57.

Chapter 5

Oscillating gene expression of the molting cycle oscillator is coordinated with pharynx growth in larvae

Joleen J. H. Traets, Timo Louisse and Jeroen S. van Zon

This chapter is a part of the following publication:

“Joleen J.H. Traets Timo Louisse and Jeroen S. van Zon. Oscillating gene expression of the molting cycle oscillator is coordinated with pharynx growth in larvae.”

(Manuscript in preparation)

5

5.1 Abstract

Oscillatory gene expression in *C. elegans* is coupled to the molting cycle, that controls the renewal of the animal's cuticle. It is an open question if oscillatory gene expression dynamics play a role in developmental processes other than molting. Interestingly, growth of the pharynx has been suggested to have a cyclical character, indicating a potential link with the molting cycle. Here, we studied the interplay between pharynx-specific oscillatory genes and the growth of the pharynx. We focused on three genes, *myo-1*, *nhr-7* and *marg-1*, belonging to a group of oscillating genes peaking in the early intermolt, at the same developmental stage at which we observed strong pharynx growth. When animals hatched without food, L1 arrested animals halted their body and pharynx growth. Moreover, we found that in those animals the pharynx genes are “locked-in” at a phase that corresponded to the expression levels around ecdysis, consistent with the observed arrest of pharynx growth. The same locked-in phase was observed in starvation-induced arrests in L3 and L4 animals. However, those animals displayed a larger pharynx, compared to body size, than seen in well-fed animals, resulting from pharynx growth arresting

later than body growth. Moreover, the time of pharynx growth arrest in L3 and L4 animals was highly variable compared to time of arrest of body growth. Overall, the observed coordination between the pharynx growth and expression oscillations of pharynx genes, both during normal development and following arrest, suggests that oscillating expression of these genes is functionally linked with organ growth.

5.2 Introduction

Oscillations are a fundamental dynamical principle in biology. From primitive bacteria to the most complex multicellular organisms, oscillations play an important role in their biological processes, such as in cell cycles, circadian rhythms, heart beat and neuronal activity (Kruse and Julicher 2005; Ebisuya and Briscoe 2018; Uriu 2016). Among the most studied biological oscillators are circadian rhythms, that are synchronized to external time, in order to adapt and anticipate on the environmental changes (Saini, Jaskolski, and Davis 2019). Oscillatory behavior in gene expression is also utilized to drive periodic events in development, without adaptation to external time, such as in the vertebrate segmentation clock (Krol et al. 2011; Kageyama et al. 2012). In development, recurring phenomena in spatial and temporal organization are common, and require re-initiation of specific periodic gene expressions programs every cycle. In the case of the vertebrate segmentation clock, the recurring gene expression programs are a chain of molecular events leading up to epithelialization of mesodermal cells for the formation of somites (Dequeant et al. 2006; Matsuda et al. 2020). Another example, is the molting cycle oscillator in *C. elegans*, that is coupled to the recurring developmental event of renewal and shedding of the cuticle (skin) of the animal during larval development (Meeuse et al. 2020; Hendriks et al. 2014; Turek and Bringmann 2014).

Whereas the output of the segmentation clock is a spatial pattern of somites, the molting cycle oscillator in *C. elegans* is intimately connected to developmental timing in the larval stages. Development of *C. elegans* larvae is divided into four distinct larval stages. Each stage starts with the intermolt, during which the animal is active and feeds, and ends with the molt, a period of lethargus during which the cuticle (or exoskeleton) of the animal is renewed and shed (Monsalve and Frand 2012; Lazetic and Fay 2017). Approximately > 20% of the transcriptome of *C. elegans* consist of genes that oscillate during development, each with large, conserved amplitudes and different peak phases (Hendriks et al. 2014; Meeuse et al. 2020). Those gene expression oscillations are coupled to molting, with the oscillations adjusting their period to the timescale of development. For example, when the temperature decreases not only the rate of development slows down, but the period of the oscillations is decreased by the same factor (Grun et al. 2014; Hendriks et al. 2014). Most of the oscillating genes are thought to be related to the molting cycle, with each gene exhibiting four peaks with the same periodicity as the molts. Also, putative cuticle components and enzymes are enriched in genes that oscillate with a similar phase as the molt itself (Hendriks et al. 2014; Turek and Bringmann 2014). However, at the same time, many genes lack a direct link with molting, and also many oscillating genes have a phase outside the molt. Hence, it remains unclear what the general function of these gene expression oscillations in larval development is.

Oscillatory gene expression occurs in different tissues of the animal (Meeuse et al. 2020). One of those tissues is the pharynx, also called the foregut. The pharynx is required for feeding, by pumping up bacteria from its external environment. The pharynx is an encapsulated structure and the lumen is lined by a pharyngeal cuticle consisting of collagen and the more stiff component chitin (Zhang et al. 2005). Similar to the cuticle of the body, the chitinous pharynx cuticle is also shed and replaced every developmental stage. Most of the development of the pharynx anatomy happens in the embryo. During larval development, the pharynx grows and it has been shown that the buccal cavity and grinder of the pharynx grow in bursts at the molt, when the old pharyngeal cuticle is released (Knight *et al.*, 2002; George-Raizen *et al.*, 2014). These observations have served as an indication for how the pharynx grows, and suggest that the pharynx growth has a cyclical character. The cyclical character of the pharynx growth might indicate a link with the oscillatory gene expression during larval development. However, a systematic dissection of pharynx growth during larval development and its relation with the molting cycle has not been performed.

Animals halt their growth and arrest their development when nutritional conditions are poor. The arrested state can be maintained from several weeks to months, after which the development can be resumed when nutritional conditions improve (Baugh 2013). Well-studied examples are the L1 arrest and dauer arrest, in which animals arrest development upon starvation, or, in the case of dauers, also crowding (Lewis and Fleming 1995; Baugh 2013). Developmental checkpoints also exist during other larval stages, such as at early L3, early L4, young adult stage (Schindler, Baugh, and Sherwood 2014; Angelo and Van Gilst 2009). Oscillating genes have been shown to arrest at a specific phase of the oscillator in larvae that are in dauer arrest (Meeuse et al. 2020; Hendriks et al. 2014; Schulenburg and Felix 2017). The specific phase at which the arrest of the oscillator occurs corresponds to the stage of development at which these developmental checkpoints occur, suggesting progression of the oscillator is tightly coupled to developmental progression. Whereas growth of the body likely halts directly upon starvation and also no longer increases when animals arrest development, it is not clear whether this is also the case for the different organs in the body, such as the pharynx. For instance, under starvation vulva development proceeds, at least in terms of cell division and changes of morphology, until the next developmental check point is reached (Schindler, Baugh, and Sherwood 2014), suggesting that under such circumstances development can be decoupled from body growth. Whether the same also holds for growth and development of other organs, for instance of the pharynx, has not been determined.

In this study, we focused on the interplay between pharynx-specific oscillating

genes and pharynx growth, both during conditions that supported growth or led to its arrest. Here, we focused on three genes, *myo-1*, *nhr-7* and *marg-1*, whose peak phase of expression in the early intermolt coincided with the period during which we observed strong growth of the pharynx. The peak of expression is coordinated with the growth of the pharynx, which grows fastest right after the ecdysis and slows down in growth half-way the developmental stages. When animals are hatched without food, they arrested body length extension and pharynx growth immediately upon entering the L1 arrest. Moreover, we showed that the pharynx genes no longer oscillate and are “locked-in” at a specific phase, in coordination with the growth of the pharynx. This locked-in phase of expression was also found in other types of more complex arrests during the L3/L4 stages of the animals upon starvation. However, those animals had a larger pharynx compared to their body, when compared to well-fed animals of the same developmental age. By following the growth of the pharynx in time while animals are starved, we showed that the growth of the pharynx is continued until around the developmental checkpoint, in contrast to the earlier halt of growth of the body length. Moreover, how pharynx growth arrested upon starvation was variable between the animals. While we measured gene expression in animals undergoing L3 or L4 arrest, our data currently cannot establish whether gene expression of those oscillating genes that are correlated with pharynx growth during normal development, also show expression peaks during continued pharynx growth under starvation. Overall, the link we observed between pharynx growth and molting cycle gene expression of pharynx-specific genes that peak during the early intermolt, suggests that expression of these genes is functionally linked with organ growth.

5.3 Results

5.3.2 Oscillating genes in the pharynx region

Here, we focus on the pharynx, the feeding organ of *C. elegans*, and the relation of oscillating gene expression with the growth of the organ. The study of Meeuse *et al.* (Meeuse *et al.* 2020), showed that oscillating gene expression occurs in different tissues including, the pharynx and hypodermis. To follow-up on the analysis of Meeuse *et al.*, we analyzed a previously published single-cell RNA-sequencing data set with tissue-specific annotation (Cao *et al.* 2017), to find genes that are expressed in the pharynx. The genes with oscillatory expression in the pharynx show phases that are widely distributed (Fig. 1A). The distribution of phases of pharyngeal genes can be divided up into five groups. To better understand the dynamics of the gene expression during larval stage progression, we tried to link the phases in the gene expression data with the different stages of the molting cycle. According to Hendrik *et al.* (Hendriks *et al.* 2014), the molt occurs between 180° and 270°, determined by

specific collagen genes peaking at the time of the molt. We also tried to estimate the phase of the ecdysis based on genes known to be related to the molt according to literature. In particular, we analyzed the phases of genes known to be related to the renewal and shedding of the cuticle (i.e., *mlt-10* and *dpy-2*) (Lazetic and Fay 2017), and APPG genes (i.e., *abu-6* and *pqn-54*), that promote the assembly and function of the pharyngeal cuticle (George-Raizen et al. 2014). Based on the phases of those genes, we estimated the approximate phase in which ecdysis occurs ($\sim 230^\circ$ - 270°), within the period of the previously reported phase of the molt (Fig. 1B, see Methods).

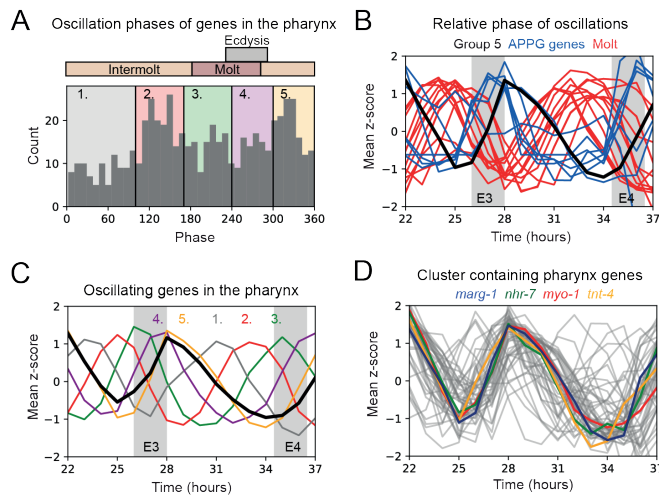


Figure 1: Pharynx-specific oscillations in gene expression. (A-D) Analysis of RNA-sequencing data on oscillatory gene expression during development in *C. elegans* (Hendriks et al. 2014). **(A)** Distribution of peak phases of genes expressed in the pharynx determined by annotated single-cell RNA-sequencing data (Cao et al. 2017). We separated genes into 5 groups of different phases (shown in different colors), based on clustering, that represent different parts of the larval stage. Top rectangles: approximate location of ecdysis, molt and intermolt. **(B)** Expression levels of molting-specific genes (red) and ecdysis-specific genes (APPG genes, blue) as function of hours after hatching. The average expression of Group 5 genes (black lines) peaks approximately right after the ecdysis (grey box), in the intermolt. **(C)** Relative timing of the average mean z-score of each grouped distribution (1: grey, 2: red, 3: green, 4: purple, 5: orange). The cluster belongs to the fifth group, consisting out of genes related to muscle contraction, defense response and actin binding based on GO term overrepresentation. **(D)** Genes in the cluster obtained with a novel percolation clustering algorithm (Chapter 4) containing pharyngeal related genes with a specific peak phase (Werner et al. 2020). All most all genes from the cluster are in group 5, and some in group 4. We selected *marg-1*, *nhr-7*, *myo-1* and *tnt-4* for in-depth analysis (colored lines), as their expression dynamics was representative of Group 5 genes. (C) The with the average mean z-score of the cluster (black) coincides with the average mean z-score of group 5.

To study the biological function of the groups in the pharynx distribution, we performed GO term overrepresentation on each group individually. GO term analysis of group 1, likely located in the intermolt (Fig. 1A-C), revealed an overrepresentation of genes related to negative regulation of endopeptidase activity. This suggests a link with the balance of protein synthesis/proteolysis in muscles during active and inactive states (Morris et al. 2005), i.e. pumping or not pumping of the pharynx. In group 2, we found an overrepresentation of peptidase activity and chitin related genes, more specifically chitinase proteins (i.e. CHT-1 and CHT-2) responsible for the breakdown of the chitin, which is found in the pharyngeal cuticle. In contrast, group 3 is overrepresented by genes that are involved chitin metabolic process related to the chitin synthesis, such as CSH-2 (Zhang et al. 2005). Genes in group 3 approximately peak during the molt, in which a new chitinous pharyngeal cuticle most likely is produced. Most genes related to assembly and function of the chitinous pharyngeal cuticle are present in group 4, at the moment of ecdysis (George-Raizen et al. 2014). Lastly, group 5 consists out of genes that have an average phase approximately right after ecdysis at the start of the intermolt. The group is overrepresented by genes involved in muscle contraction and defense response against bacteria. Moreover, the group consists several structural genes required for the functioning of the pharyngeal muscles, such as, myosins and troponins (i.e. *myo-1*, *myo-2*, *ttn-4* and *tnc-2*), suggesting a potential link with growth of the pharynx.

In the previous chapter (**Chapter 4**), we found a cluster of pharyngeal genes in RNA-sequencing data of young adult animals with a novel percolation clustering algorithm that showed correlated variability in expression between individual animals (Werner et al. 2020). Interestingly, almost all genes from the pharynx cluster are in group 5 (Fig. 1C-D). Both the cluster genes and group 5, contain genes that are structural components of the pharyngeal muscles. This observation provides a potential explanation for the emergence of this cluster and might indicate that these genes peak in expression also in the young adults we isolated for RNA-sequencing analysis, who just emerged from the L4 larval stage.

5.3.2 Oscillating expression levels of genes expressed in different pharyngeal cell types

To study the spatial and temporal expression pattern of the oscillating genes in the pharynx more in depth, we used single molecule FISH (smFISH) to quantify, more precisely than RNA-sequencing methods, mRNA molecules in the pharynx region (Raj et al. 2008). We selected three oscillating genes from the group of pharyngeal genes with a potential relation to growth, *nhr-7*, *marg-1* and *myo-1* (Fig. 1D), based

on the previously stated requirements for smFISH experiments in **Chapter 4**. Other requirements are: expression in different tissues of the pharynx (e.g. muscle, marginal or neuronal cells), and a combination of structural and non-structural genes. Of the three genes, *myo-1* is a structural component of the pharyngeal muscles, whereas the other two, *marg-1* and *nhr-7* are non-structural. The nuclear hormone receptor *nhr-7* is a transcription factor with unknown function. Null-mutants of *nhr-7* show normal pharynx development on food, and arrest of pharynx development when in L1 arrest (Fig. S1A-B). Even though the precise function of *marg-1* is not known, *marg-1* is specifically expressed in the marginal cells of the pharynx, that are interspersed between the pharyngeal muscle cells and are responsible for scavenging cholesterol and other hydrophobic small molecules during feeding (Kamal et al. 2019). The myosin heavy chain *myo-1* (MHC D) is known to be exclusively expressed in the pharyngeal muscles where, as a structural constituent of the muscles, it is required for feeding (Ardizzi and Epstein 1987; Miller, Stockdale, and Karn 1986).

We confirmed that all three genes are expressed in the pharynx region. More specially, *nhr-7* and *myo-1* are expressed in the pharyngeal muscle cells (pm3-7 cells) and *marg-1* is expressed in the marginal cells (mc1-3 cells) (Fig. 2A). Notably, *nhr-7* is also expressed in the amphid sheath cells (based on *vap-1* expression) and in unknown cells with a similar cell shape. Because the whole pharynx in animals staged older than L3 does not fit in the field of view of the microscope, we quantified the expression levels within a smaller region as representative for the whole pharynx. This region consists out of the so-called isthmus region, more specially, the pm6 and pm5 cells for *nhr-7* and *myo-1*, and mc2 cells for *marg-1*. When we estimated developmental age by body length of individual animals, we confirmed that the expression of all three genes oscillates throughout larval development. All three oscillating genes show a strong peak of expression right after the ecdysis, at the beginning of the L2-L4 larval stages (Fig. 2A-D), but we did not observe a peak of expression of any of the three genes in the L1 larval stage. We found that high *mlt-10::GFP::pest* expression, a marker of lethargus (Meli et al. 2010; Frand, Russel, and Ruvkun 2005), correlated with low *nhr-7* expression (Fig. S1C), confirming that the peak of expression occurred not during the molt, but after ecdysis. Whereas *nhr-7* and *marg-1* expression almost vanished in the isthmus, *myo-1* expression did not show a pronounced decrease (Fig. 2B-C), likely reflecting its requirement for continuous expression as a structural gene required for muscle function. However, we still observed peaks in *myo-1* expression right after ecdysis.

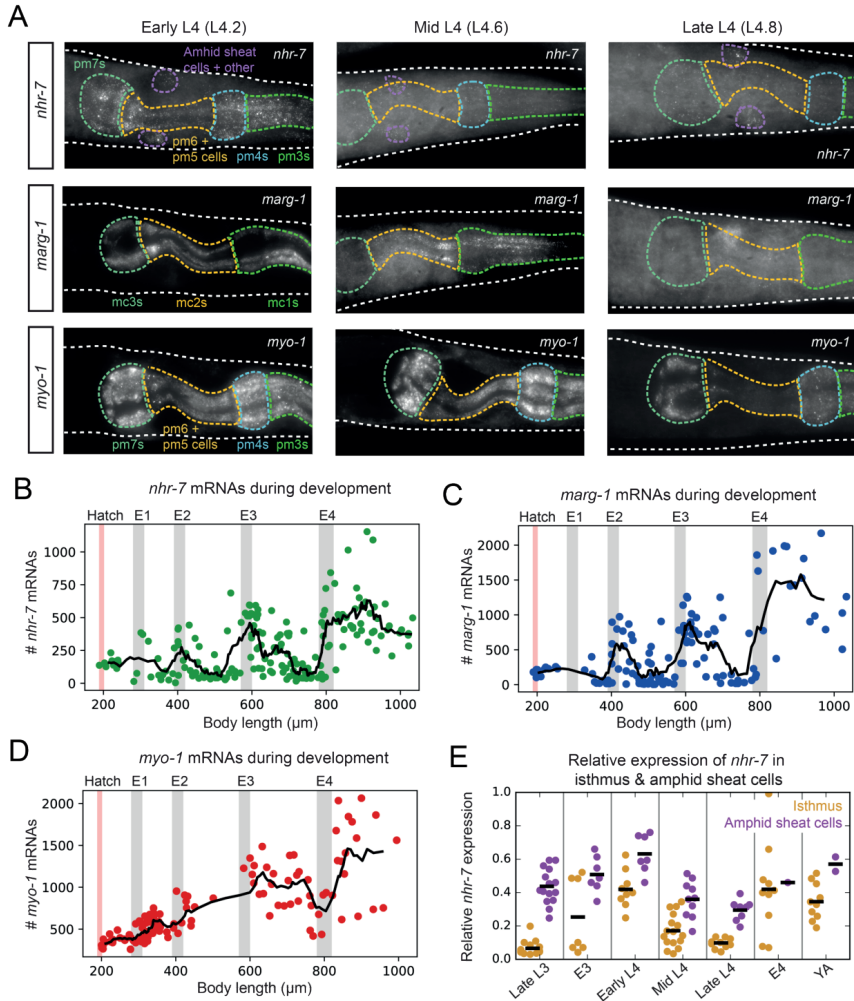


Figure 2: Oscillations in expression of pharynx genes peak directly after ecdysis. (A) smFISH microscopy images of the three selected oscillating genes in the pharynx, staged based on vulva and gonad morphology. *nhr-7* and *myo-1* are expressed in the pharyngeal muscle cells (pm3-7), while *nhr-7* is also expressed in amphid sheat cells and other unidentified cells. *marg-1* is expressed in the marginal cells in the pharynx (mc1-3). Coloured regions depict different cell types as indicated. Orange region represents region that is used for quantification. (B-D) Expression levels of *nhr-7*, *marg-1* and *myo-1* in the isthmus region as function of body length. All three genes exhibit oscillations in expression with a peak directly after ecdysis. Grey insets represent the approximate time of the ecdysis, estimated by body length. Black line is the rolling mean. (E) *nhr-7* expression dynamics in the isthmus region (green) and amphid sheat cells (purple) for subsequent developmental stages (late L3, M3, early L4, mid L4, late L4, M4 and YA, based on vulva, gonad and sperm morphology). Oscillations of *nhr-7* expression exhibit clear amplitude difference between these cells.

Next, we examined whether the phase of the oscillations of the genes differed between the different cell types in which the genes are expressed. For example, *nhr-7* is expressed in the pharyngeal muscle but also the amphid sheath cells located outside the pharynx region. Here, body length as an indication for the developmental stage did not provide enough time resolution. Instead, we determined the developmental stage based on the gonad, vulva and sperm morphology (Mok, Sternberg, and Inoue 2015; Killian and Hubbard 2005; Kimble and White 1981). Indeed, by subdividing the L4 stage in 10 sub-stages based on vulva morphology, we found differences in timing of *marg-1* expression between the marginal cells. During the expression peak right after the ecdysis of *marg-1*, first the mc1 cells start expressing high levels of *marg-1* at the L4.1 sub-stage, after which the mc2-3 cells follow at the L4.2 sub-stage (Fig. 2A). To study the differences in oscillation phase of *nhr-7* between the pharynx and different cell types located outside the pharynx, we grouped animals into the following sub-stages: prior to L3 molt, L3 molt, early L4, mid L4, late L4, L4 molt and young adults (YA). The phase of the oscillations of *nhr-7* in the pharynx compared to the amphid sheath cells differed only slightly. However, the amplitude is higher in the amphid sheath cells (Fig. 2E).

To summarize, oscillatory expression of the three pharynx genes, *nhr-7*, *marg-1* and *myo-1* peak right after ecdysis. Expression levels of *nhr-7* and *marg-1* decrease more than *myo-1* during the molt, likely due to its structural function in the pharyngeal muscles. The oscillation phase and amplitude of the genes was shown to differ between cells, indicating that the regulation of the oscillations depends on cell identity.

5.3.2 Growth of the pharynx is coordinated with oscillatory gene expression

Even though the identity of the genes in Group 5, that peak in expression in the pharynx at the start of the intermolt, appears related to growth, it is not known how growth of the pharynx is regulated in time. Here, we used time-lapse microscopy (Gritti et al. 2016) to directly follow the growth of the pharynx in time in individual animals throughout larval development (Fig. 3A). The pharynx can be subdivided into six sections, from anterior to posterior: the buccal cavity, procorpus, metacorpus, isthmus, terminal bulb and pharyngeal-intestinal valve. First, we measured the length of the entire pharynx (start buccal cavity until pharyngeal-intestinal valve), as a measure of growth of the entire organ based on transmitted light images, from hatching to adulthood (Fig. 3B). In contrast to the previously shown step-wise growth of the buccal cavity and grinder (George-Raizen et al. 2014; Knight et al. 2002), our measurements revealed that the pharynx length increases continuously, but with intervals of slower growth (Fig. 3B,D). Generally, the growth of the pharynx is fastest right after the ecdysis and becomes slower half-way towards

the next ecdysis, a phenomenon most pronounced in the L2-L4 stage. In contrast, during the L1 stage the growth of the pharynx appears almost linear. In some cases, faster growth of the pharynx is already visible right before ecdysis.

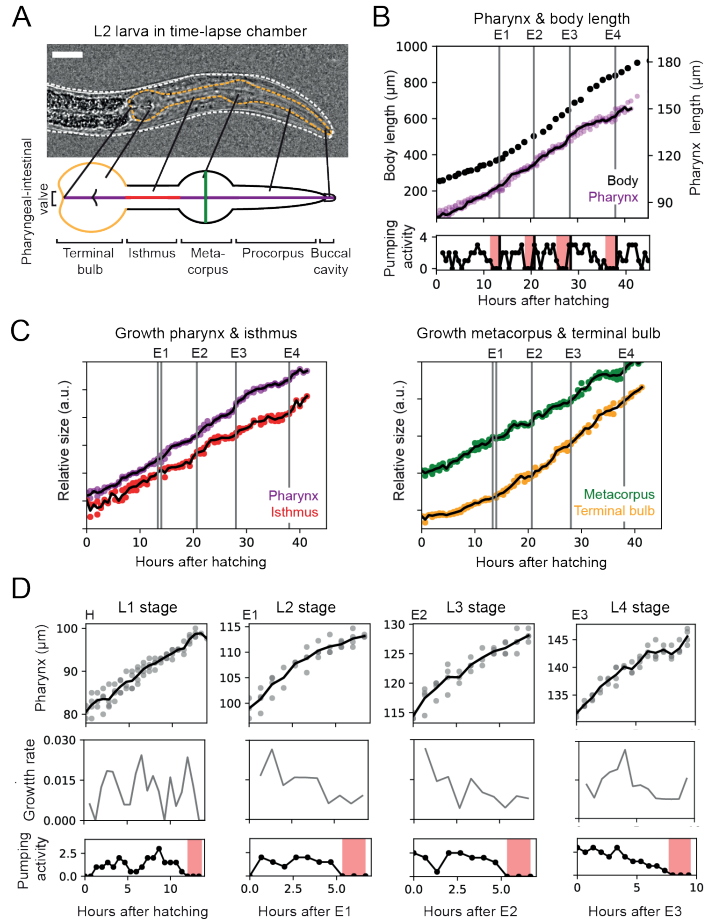


Figure 3: Non-uniform pharynx growth as function of larval stage progression. (A) The pharynx is divided into six sections, the buccal cavity, procorpus, metacarpus, isthmus, terminal bulb and pharyngeal-intestinal valve. All sections can be measured during development in the time-lapse setup (Gritti et al. 2016). Color in the schematic corresponds to the measures of pharynx growth in panels (B),(C). White: body outline, yellow: pharynx outline. Scale bar 15 μm. (B) Growth of body (black) and pharynx length (purple) in one representative animal imaged from hatching to adulthood. In contrast to the increase of body length, pharynx length shows intervals of slower growth in the latter half of each larval stage apart from L1. Bottom panel shows the pumping activity of the pharynx in time. Red inset: approximate time of lethargus is determined from pumping activity, which is reduced during the lethargus. (C) Growth of different pharynx subsections (n=2) show similar behavior as the total pharynx length, with longer, more pronounced periods of slow

growth. Purple: pharynx length. Red: Isthmus length. Green: Metacarpus width. Orange: Terminal bulb area. **(D)** Pharynx length (top panels), pumping activity (middle panels), and pharynx growth rate (bottom panels) as function of time for the L1-L4 larval stages (n=4). Growth rates decrease already before the lethargus, as determined by average pumping activity

Instead of measuring the entire pharynx represented by the length of the pharynx, we measured the individual sections of the pharynx in time, to study possible differences in growth (n=2, Fig. 3A). First, the length of the isthmus follows the same growth trend as the length of the pharynx (Fig. 3C). The growth of the metacarpus, measured as the width of the widest part of the metacarpus, appears more linear in time, with exception of the L4 stage. Last, the terminal bulb grows similarly as the pharynx length. However, the period of fast growth at start of the intermolt seems longer, with a shorter period of slow growth during the molt. Taken together, the different sections show some slight differences between each other in growth, but all measured sections maintain a similar growth trend as the pharynx length, with an increase in growth right after ecdysis and a decrease in growth towards the next ecdysis.

The time of slower growth of the pharynx appeared similar to that of the lethargus, which occurs prior to the ecdysis, during which animal remain quiescent and stop pharyngeal pumping (Raizen et al. 2008; Singh and Sulston 1978). To confirm this, we quantified the average pumping activity as an indication for lethargus, by counting the number of pumps within one image stack. Indeed, we found that the periods of decreased pumping activity coincided with the slower growth of the pharynx (Fig. 3B,D). However, the pharynx already slows down in growth before the lethargus, suggesting that the growth is reduced not purely by the lack of food uptake.

As shown previously, the peak of expression levels of the oscillating genes *nhr-7*, *marg-1* and *myo-1* happens directly after the ecdysis and the lowest expression levels are found during lethargus. Similarly, pharynx growth is fastest after the ecdysis and starts declining half-way the larval stages, but prior to the lethargus. Gene expression oscillations are absent and also seemingly arrested in well-fed early L1 animals according to the study of Meeuse *et al.* (Meeuse et al. 2020), which is in line with our limited observations on expression levels of the three pharynx genes during the early L1 stage (Fig. 2B-E). Interestingly, compared to later larval stages, we found that the growth of the pharynx in L1 was linear. Overall, our observations show that the growth of the pharynx peaks together with the expression levels of the Group 5 oscillating genes in the pharynx, suggesting that they are coordinated in time.

5.3.3 Arrested L1 larvae halt pharynx growth and gene expression oscillations

The oscillating expression levels of *nhr-7*, *marg-1* and *myo-1* are coordinated with the growth of the pharynx during development under well-fed conditions. To study the effect of arrested growth on the expression levels of pharynx genes, we hatched larvae without food to induce L1 arrest, the well-described developmental arrest state (Baugh 2013). L1 arrested animals are stress resistant and can remain in L1 arrest for several weeks (Lewis and Fleming 1995). Moreover, in order to achieve such resistance, during L1 arrest growth is halted and the metabolism is altered in order to survive prolonged starvation (Kostal 2006).

When larvae hatch in absence of food, by placing bleached eggs in time-lapse chambers without food, the animals indeed halt the growth in body length immediately (Fig. 4A-B). The pharynx growth is also immediately arrested after hatching in the absence of food. Even after 20 hours of starvation, L1 arrested animals have a similar pharynx and body length compared to animals right after hatch (Fig. 4B). Moreover, the terminal bulb and length of the isthmus also remains stable in size during prolonged starvation of 20 hours (Fig. S4).

To study if the gene expression oscillations in *nhr-7*, *marg-1* and *myo-1* also arrest, we hatched animals in absence of food and quantified their expression levels at different time points. All three genes show decreased expression levels after 5 hours and 18 hours of starvation compared to expression levels of well-fed animals (Fig. 4C), even though some animals still show relatively high *nhr-7* expression levels after 5 hours of starvation. In a recent study of Meeuse *et al.* (Meeuse *et al.* 2020), they showed that gene expression oscillations arrest at a specific oscillator phase both in larvae that arrest as dauers and in adult animals that no longer molt. Moreover, within the limits of their RNA-sequencing data resolution, the specific phase corresponds to a phase around ecdysis. Here, the low expression levels of *nhr-7*, *marg-1* and *myo-1* after 18 hours of starvation during L1 arrest are consistent with the recent observations of a “locked-in phase” of the oscillator, in which the expression of the oscillating genes are stuck in a phase around the ecdysis.

Taken together, L1 arrested animals halt growth of pharynx and body immediately after the hatch, and the pharynx does not change in size during prolonged starvation (20 hours). The oscillating genes *nhr-7*, *marg-1* and *myo-1* no longer show oscillating behavior during L1 arrest, and are locked-in in a specific phase of the oscillations which corresponds to a phase around ecdysis. The lower expression levels of the pharynx genes in L1 arrested animals that halt growth of the pharynx, again indicates coordination between gene expression oscillations and the growth of the pharynx.

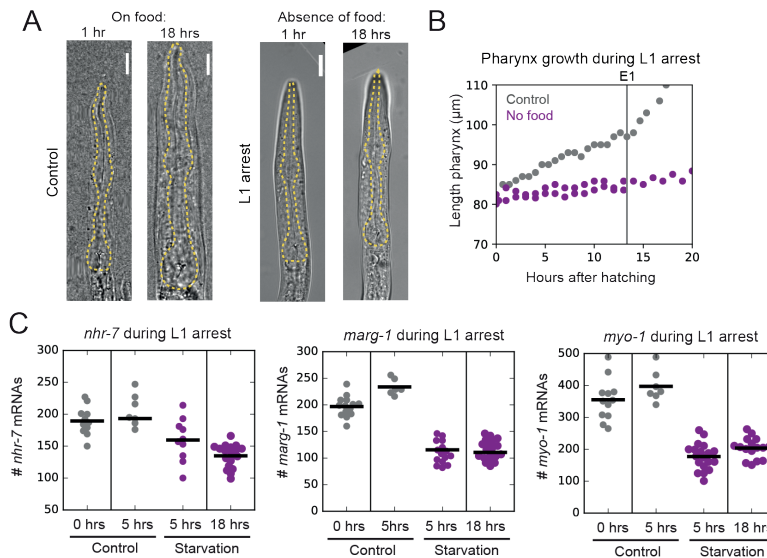


Figure 4: L1 arrested animals halt pharynx growth and gene expression oscillations. (A) Individual animals imaged by time-lapse microscopy on food or in absence of food, imaged 1 hr. and 18 hrs. after hatching. L1 arrested animals maintain a similar-sized pharynx after 18 hrs. without food, while animals on food show growth of the pharynx. Scale bar 15mm. **(B)** Growth of pharynx length in a representative animal on food (grey) and in animals (n=2, purple) without food. **(C)** Expression levels of *nhr-7*, *marg-1* and *myo-1* in L1 arrested animals after 5 hrs. and 18 hrs. in absence of food, compared to control animals on food right after hatch and after 5 hours of development. All three genes have decreased expression levels and show no indication of oscillatory behavior in L1 arrested animals.

5.3.4 Arrested gene expression oscillations in late-larval developmental arrest

So far, we observed that the expression levels of *nhr-7*, *marg-1* and *myo-1* in L1 arrested animals are locked-in at a phase corresponding to a phase around ecdysis. The population of L1 arrested animals all experienced starvation at the same developmental stage, when oscillations in gene expression are still absent, meaning that all animals had similar expression levels at the start of starvation. It was recently shown that larvae at later developmental stages can arrest under starvation at specific developmental checkpoints similar to the L1 arrest (Schindler, Baugh, and Sherwood 2014). In contrast to the L1 arrest, larvae that are exposed to starvation at later larval stages already exhibit oscillatory expression of pharynx genes. Hence, gene expression levels will vary depending on the exact time animals are exposed to starvation. Therefore, we examined how oscillating gene expression responds to such late-larval developmental arrest.

When animals aged between L3-L4 stage are starved, the animals can go into one of the four types of arrest/diapause; L4 arrest, young adult arrest, adult reproductive diapause (ARD) and “bagged” fate, depending on the developmental stage and the density of the population (Schindler, Baugh, and Sherwood 2014; Angelo and Van Gilst 2009; Chen and Caswell-Chen 2003) (Fig. 5A-B). In contrast to the study of Schindler *et al.*, we included the ARD stage and renamed the L4 and young adult arrest to L3 and L4 arrest, respectively, in line with the Angelo *et al.*, study. We starved L3-L4 larvae for 24 hours to induce developmental arrest and then quantified the expression levels of the oscillating genes with smFISH. Indeed, we found that after 24 hours of starvation, all animals were in one of the four arrested states. As previously, we determined the developmental stage of the arrested and well-fed animals based on the gonad and vulva morphology, and on the presence or absence of sperm and eggs. The control animals for L3 and L4 arrest show high variability, reflecting the peak in oscillations in the early L3 and L4 stage. In contrast, animals in L3 or L4 arrest show for all three genes decreased expression levels with less variability, indicating arrest of the oscillations (Fig. 5C-E, $p < 0.05$). Animals in ARD have lower expression levels than well-fed control animals at the same stage ($p < 0.05$). Also, animals in ARD have slightly lower expression levels compared to L4 arrested animals ($p < 0.05$), possibly because arrest and diapauses represent states that differ substantially, e.g. the ability to withstand prolonged periods of starvation (Kostal 2006; Baugh 2013). Interestingly, bagged animals have similar *nhr-7* expression levels when compared to well-fed control animals, suggesting that bagged animals do not respond to starvation with changes in the expression levels of *nhr-7*.

Similar to the L1 arrest, the expression of the oscillating genes, *nhr-7*, *marg1* and *myo-1* in animals arrested in L3 arrest, L4 arrest and ARD is locked-in at a specific phase of the oscillations, despite the added complexity of the arrests. Our observations on expression levels of the pharynx genes in the more complex arrest stages, are in line with the previously observed locked-in phase in L1 arrested larvae, dauer larvae and adult animals. Moreover, the locked-in phase of the more complex arrest stages corresponds to either the time of ecdysis before the peak or to the early L4 stage/pre-young adult stage right after the peak (Fig. 2B-E).

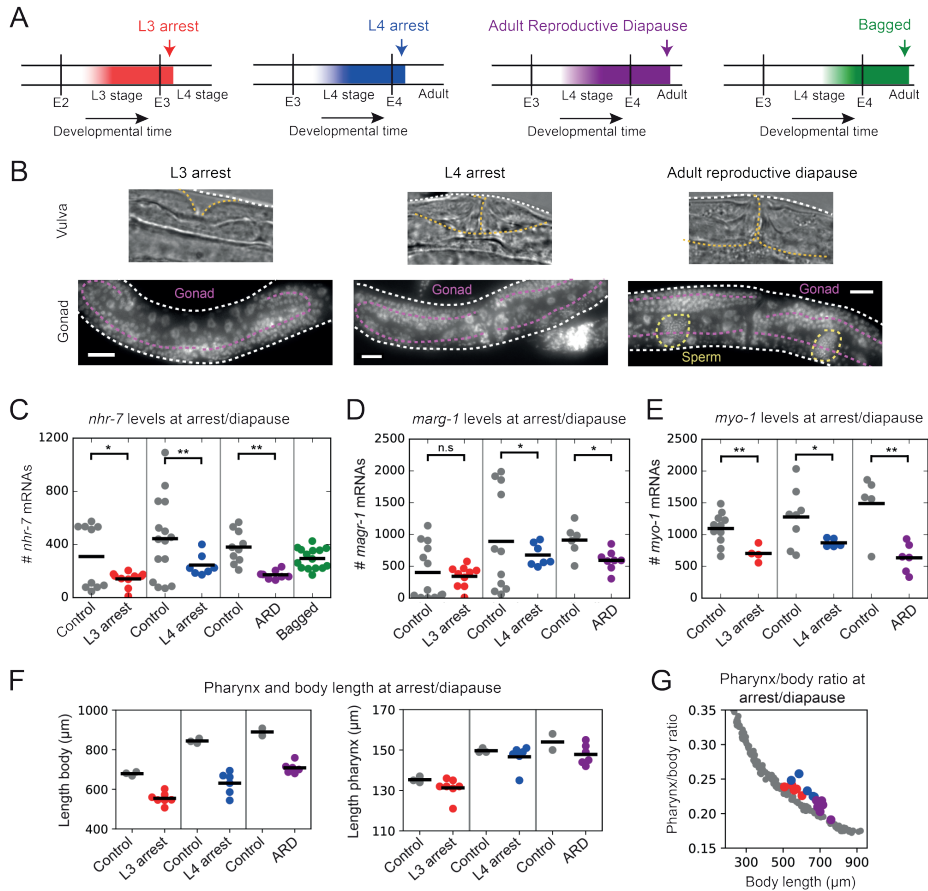


Figure 5: Dynamics of gene expression oscillations and growth in late-larval developmental arrests. (A) Schematic overview of late-larval arrests: L3 arrest, L4 arrest, adult reproductive diapause (ARD) and 'bagged' fate (egg retention), with entry of arrest depending on developmental stage at time of starvation, as indicated schematically by the color gradient. (B) Morphology of vulva and gonad of animals in L3 arrest, L4 arrest and ARD, determined with intermittent light (top) and DAPI signal (bottom). L3 arrest: L4.1 vulva morphology and arrest gonad arm extension. L4 arrest: L4.9 vulva and fully extended gonad arms. ARD: spermatogenesis. Scale bar 20mm. (C-E) Expression levels of *nhr-7*, *marg-1* and *myo-1* in L3-arrested (red), L4-arrested (blue), ARD (purple) and bagged animals (green), compared to well-fed animals at similar developmental stages, as determined by vulva/gonad/sperm morphology. All three genes have lower average expression levels compared to respective controls. Well-fed control animals around L3 and L4 molt either expressed very low or very high levels of *nhr-7* and *marg-1*, reflecting the rapid increase in expression at this stage under normal conditions. This was absent in arrested animals. * $P < 0.05$ and ** $P < 0.01$ (Welch's t-test, with Levene's Test for equality of variances). (F) Pharynx and body length in animals at L3 arrest (red), L4 arrest (blue) and ARD (purple), compared to well-fed animals. Body length is significantly

shorter in arrested animals, while pharynx lengths are more similar to the pharynx in well-fed control animals. **(G)** Ratio between pharynx and body length in arrested animals, L3 arrest (red), L4 arrest (blue) and ARD (purple). In well-fed animals, the size of the pharynx decreases compared to body length (grey). However, most arrested animals have a larger pharynx relative to body length, compared to well-fed animals.

5.3.5 Variation in pharynx/body size ratio in late-larval arrests

In L1 arrested animals, we found that the pharynx halted growth of pharynx and body immediately upon hatching. To study the differences in body and pharynx size in animals in the more complex arrested states, we starved the animals in the L3-L4 stage again for 24 hours. All arrested animals have a consistently shorter body length compared to the well-fed control animals with a similar developmental stage (Fig. 5F), indicating that body length extension decreased already before entering developmental arrest. However, the length of the pharynx in arrested animals is more variable compared to the well-fed control. When we calculated the ratio of pharynx and body length as indication for the relative pharynx size, we found that most arrested animals have a larger pharynx relative to their body (Fig. 5G), a phenotype that was most pronounced in animals in L4 arrest and ARD.

These results suggested that the arrest of pharynx growth is more complex than the direct halt of growth of the body length upon starvation. At the same time, oscillatory gene expression was consistently locked-in at the same phase in each of the arrested stages, as also seen in L1 arrested animals. However, in contrast to the L1 arrested animals, in late-larval arrest pharynx growth did not simply halt with the arrest in body length extension, raising the question whether pharynx growth and oscillatory gene expression are uncoupled for these arrests. However, the data on arrested animals does not provide information on the growth of the pharynx or the gene expression levels as function of time while animals are proceeding towards the arrest.

5.3.6 Variability in arrest of pharynx growth in starving animals

In arrested animals the pharynx/body ratio was higher compared to control animals in the same developmental stage, meaning that the pharynx is slightly bigger in those animals compared to their corresponding body size. Moreover, arrested animals, generally, have a similar sized pharynx compared to well-fed animals. These observations suggested that the growth of the pharynx continues upon starvation, while the body length extension halts earlier.

To examine how this difference between pharynx and body size emerges in time, we used time-lapse microscopy to follow developmental progress of animals as they exhausted the limited food supply present in the microchambers. In this manner, we measured the growth of the pharynx and body while animals progressed into L3 arrest, L4 arrest/ARD (between which we cannot distinguish because sperm morphology is not visible with the current set-up). Indeed, we found that the increase of body length halts earlier than the pharynx length (Fig. 6A). In addition, most animals had a slightly larger pharynx compared to the body at the end of the time-lapse experiment (Fig. S3). This confirms that upon starvation growth of the body is uncoupled from pharynx growth. Interestingly, some animals kept growing their pharynx with the same rate as well-fed animals, while others halted their pharynx growth earlier and continued the growth later in time. This shows that there is a strong variability in the manner at which the animals arrest their pharynx growth upon starvation.

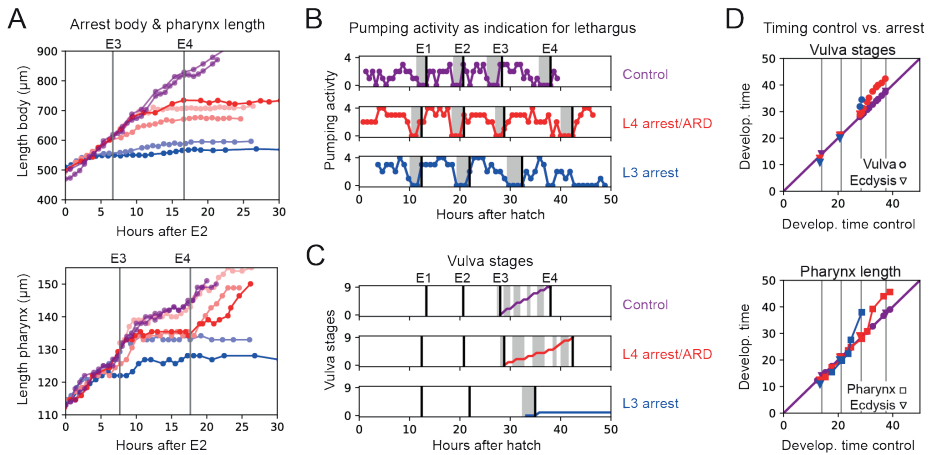


Figure 6: Variability in arrest of pharynx growth in animals upon starvation. (A) Animals arresting in time by depletion of food in time-lapse chambers. The body (top panel) and pharynx (bottom panel) length were measured in animals progressing towards L3 arrest, L4 arrest/adult reproductive diapause (ARD). Body length growth is halted earlier than pharynx growth in arrested animals. The pharynx growth arrests either immediately, following the control growth curve of well-fed animals, or with a ~2h delay. In some animals, pharynx growth paused, resuming after an ~10hr delay. Vertical black lines: average ecdysis timing well-fed animals. Purple lines: well-fed animals (n=2). Blue shaded lines: L3 arrest. Red shaded lines: L4 arrest/ARD. (B) Pumping activity and (C) vulva development in L3 arrested (blue) and L4 arrested/ARD (red) animals in animals that paused pharynx growth. Pumping activity was used as indication for lethargus (grey), while vulva development was divided in 10 stages, as determined by vulva morphology. Both molting cycle progression (B) and vulva development (C) progressed more slowly prior to entering developmental arrest. (D) The developmental timing of the vulva development in the L4 stage (top panel), and the pharynx growth (bottom panel) is compared in arresting animals (L3 arrest: blue, L4 arrest: red) versus well-fed animals (grey). Development of both vulva and pharynx in arresting animals show a

short delay in timing, after which the development is resumed at a rate similar to well-fed animals. Round: vulva stages. Square: pharynx length. Triangle: ecdysis.

While some animals going into L3 arrest and L4 arrest/ARD follow the same rate of pharyngeal growth as in well-fed animals, others seem to transiently slow down their growth. We wondered if this transient slowdown in growth is specific for the pharynx in those animals, or if other developmental processes are also slowed down in a similar manner. Using lack of pumping activity as indication of lethargus, we found that lethargus was also delayed in animals that progressed towards arrest with slow pharynx growth (Fig. 6B). We also examined the rate of vulva development during the L4 stage (Mok, Sternberg, and Inoue 2015). With the aid of vulva morphology, the L4 stage is divided up into 10 defined vulva stages (Fig. 6C). Similar to the transient slowdown of pharynx growth, the vulva development progression also occurs at a slower pace in the same animals. Moreover, the last vulva stage (L4.10) was reached at the correct time relative to the larval stage, i.e. at the end of the lethargus. To find out if the rate of the progression of the vulva stages is also similar to the rate of pharynx growth, we plotted the developmental timing of the vulva and pharynx growth in well-fed control versus animals with slowed down progression into L3 arrest and L4 arrest/ARD (Fig. 6D). Both the progression of the vulva stage and growth of the pharynx, have an initial decrease in the progression rate, but after ~5 hours the progression rate increases to levels similar to well-fed animals. Overall, these results showed that transient slow pharynx growth was accompanied by an equal slowdown in of vulva developmental and molting. Our observations raise the question whether the expression oscillations pharynx genes also continue under these starvation conditions, coordinated with pharynx growth.

5.3.7 Variability in gene expression levels of oscillating genes in animals upon starvation

Animals going into arrest upon starvation showed a high variability between the growth of the pharynx when animals going into arrest. In some animals the pharynx growth was similar to well-fed animals and the growth arrested around the developmental checkpoint. While, in other animals the progression of pharynx growth was slowed down together with other developmental processes. These observations suggest that if gene expression of the pharynx genes is coordinated with pharynx growth, then oscillations are expected to continue upon starvation until the developmental checkpoint is reached. To explore if the oscillations in the gene expression of the pharynx genes continue, we followed the expression of *nhr-7*, *marg-1* and *myo-1* at two time points, 1.5 hours and 5 hours after initiation of starvation in a population of L3/L4 staged animals. Since growth of the body is halted rapidly upon starvation, we used the vulva/gonad/sperm morphology rather

than body length to determine the developmental stage of the animals.

There are two possible scenarios on what could happen with the gene expression oscillations in the pharynx upon starvation: in the first scenario, the oscillations are immediately arrested upon starvation induction, meaning that the arrested phase depends on the time of starvation induction. In the second scenario, the oscillations in gene expression are continued. The immediate arrest of oscillations in gene expression is already excluded by the long-term starvation data, since arrested animals show a specific arrested phase of the oscillations (Fig. 5C). Subsequently, we expect that the oscillations in gene expression in the pharynx continue until the developmental checkpoint is reached.

L3/L4 staged animals subjected to immediate starvation for 1.5 hours show decreased average expression levels of *nhr-7* and *marg-1* in the L4 stage, while in *myo-1* a decrease of expression levels is less pronounced in most developmental stages (Fig. 7A). Moreover, animals that are subjected to immediate starvation for 5 hours, show a slighter decrease in expression levels of *nhr-7* and *marg-1* in the L4 stage (Fig. 7B). In most cases, the pharynx genes do not show the same gene expression levels as arrested animals when compared to the corresponding developmental stage (Fig. 7C). Moreover, this difference suggests that the locked-in phase have not been reached yet in all animals after 5 hours of starvation. Taken together, these observations hint towards continued oscillations, since 1.5- and 5-hours of starvation retain the trend of oscillatory gene expression behavior. Albeit, the expression levels of *nhr-7* and *marg-1* are decreased and *myo-1* hardly changed, suggesting that the pharynx genes do not react homogenously to starvation and perhaps some genes change amplitude.

A challenge in the interpretation of Fig. 7 is that smFISH data only provides a snapshot of gene expression in each animal in time, while our time-lapse data on pharynx growth during starvation (Fig. 6A) indicated that the progress of development was highly variable between animals. To directly link continued pharynx growth after starvation to continued oscillatory expression of pharynx genes, it is necessary to image gene expression directly in live animals, using fluorescent transcriptional reporters. While work on construction of such reporters is ongoing, we were not able to perform such experiments at this time. Based on our smFISH experiments in Fig. 7, we predict that animals that resume pharynx growth after a transient slowdown, will also show a peak of pharynx gene expression that is delayed by the same amount of time.

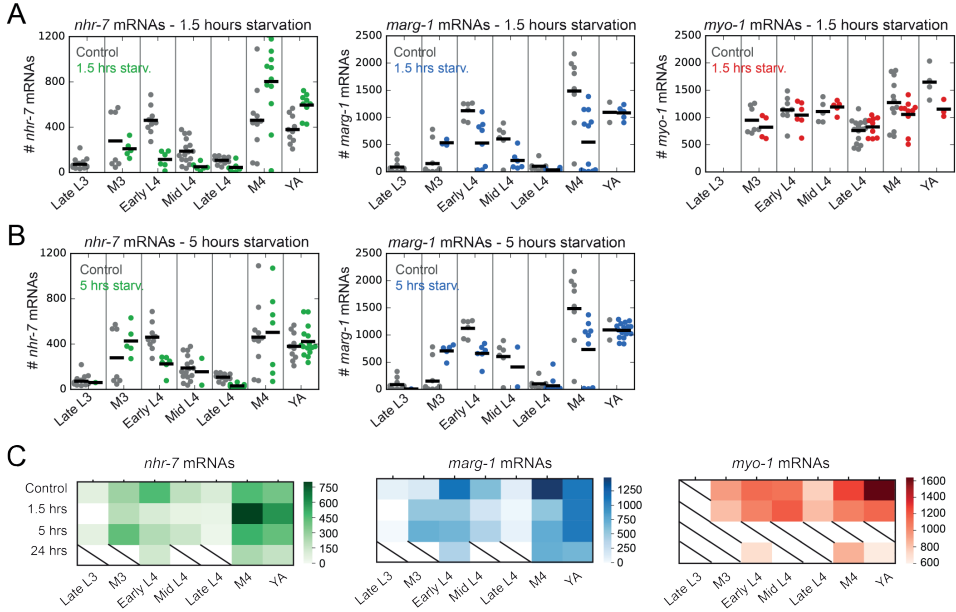


Figure 7: Changes in gene expression levels of oscillating genes upon late-larval starvation. (A) Expression levels of *nhr-7* (green), *marg-1* (blue) and *myo-1* (red) after 1.5 hours of starvation at L3-L4 stages. Both *nhr-7* and *marg-1* show a decrease in expression levels compared to well-fed control, while *myo-1* shows in some stages a decrease in expression levels. **(B)** Same as in (A), but with 5 hours of starvation instead of 1.5 hours starting at L3-L4 stages. Both *nhr-7* and *marg-1* show a slightly decrease in expression levels compared to well-fed control. The overall trend of oscillatory behavior, with only a slight difference between the well-fed animals and animals after 5 hours of starvation, suggests that the oscillations are continued. **(C)** Heatmaps of expression levels of *nhr-7*, *marg-1* and *myo-1* as function of developmental stage, comparing well-fed control, 1.5 hours of starvation, 5 hours of starvation and arrested animals. Combined data from Fig. 7A-B and 5C-E.

5.4 Discussion

In this study, we explored the relation between oscillating gene expression of pharynx-specific genes and the growth of the pharynx. Using previously published data sets, we identified a group of oscillating pharyngeal genes that oscillate in a phase at the start of the intermolt and are overrepresented by genes that appear related to pharynx growth. When we examined pharynx growth in time in individual animals, we found that the pharynx increased in size in a cyclical manner, with growth fastest directly after ecdysis and slowing down in the second half of the larval stage. Moreover, the growth appeared to be coordinated in time with the expression of the pharyngeal genes, as their oscillations peaked at the same stage of the intermolt.

We found that this coordination between pharynx growth and oscillatory expression of pharynx genes extended to arrest of pharynx growth under starvation-induced developmental arrest. L1 arrested animals halted both pharynx growth and expression oscillations, with the latter “locked-in” at a phase corresponding to a time around ecdysis. Animals undergoing late-larval arrest showed a similar locked-in phase of the oscillating pharynx genes. It is an important open question if oscillating gene expression causes growth of the pharynx, e.g. by increasing levels of structural proteins such as *myo-1* and *myo-2*, or whether growth is controlled by other mechanisms, and the peak of pharynx-specific gene expression only coincides with growth to supply the necessary (structural) components required for the larger volume of the organ. Differentiating between these two scenarios would likely require abolishing or shifting the peak in pharynx-specific gene expression and examining whether this halt or shifts pharynx growth accordingly. However, this requires knowledge of the mechanisms and genes responsible for inducing the oscillatory behavior of the pharynx-specific genes, which are currently unknown.

In the time-lapse experiments, we found that arrest of pharynx growth was more variable and occurred well after the arrest of growth of the body. We speculate that this variability in (arrest of) pharynx growth might be related to the exact developmental stage at which starvation is initiated. Moreover, the developmental stage of an animal could potentially lead to a specific response to starvation, as we have observed for the dependence on developmental stage in the starvation response of *daf-7* in Chapter 3. However, in our set-up, with chambers running out of food, the exact time at which animals experience starvation could not be measured. By providing direct control over food supply, time-lapse imaging using a microfluidics set-up could provide more information on the link between immediate starvation response and the variability in pharynx arrest (Keil et al. 2017).

The oscillator related to the molting cycle is proposed to be operating near a Saddle Node on Invariant Cycle (SNIC) bifurcation, implying that oscillations are continued until the oscillator arrests at the developmental checkpoint (Meeuse et al. 2020). This arrest of the oscillator at the developmental checkpoint agrees with our observations in L1 arrested animals, in which the gene expression oscillations halt at a phase corresponding to a time around ecdysis, the same stage as in developmental checkpoints. The arrest of oscillations in late-larval stage arrests had not been tested before. Our smFISH data on arrested animals in late-larval stages showed similar results, with oscillations arrested in the ecdysis phase. However, the smFISH data on animals that were going into developmental arrest, revealed more complex dynamics in the expression levels. Specifically, the expression levels of *nhr-7* and *marg-1* decreased upon immediate starvation of 1.5 and 5 hours of starvation

compared to well-fed animals of similar developmental age, suggesting that the oscillations are perhaps continued at a lower amplitude. Multiple genes are known to change expression upon starvation (Harvald et al. 2017). Our results suggest that oscillatory gene expression in the pharynx is also affected by the starvation response. Moreover, there are examples in which the oscillations in gene expression are dampened, for example the pulsatile dynamics in *lin-14* are dampened by the negative regulator *lin-4* miRNA (Kim, Grun, and van Oudenaarden 2013). Taken together, the oscillations in gene expression in the pharynx are most likely continued upon induction of starvation until the developmental checkpoint is reached, and our observations suggest that the oscillatory behavior of some genes are modulated upon starvation.

Our results demonstrate a precisely regulated timing of pharynx growth. The growth rate of the pharynx is highest during the intermolt when the animal is active and pumping for food, and slows down halfway through the larval stage. It is an open question why the pharynx growth happens specifically during the early intermolt, and ceases during the molt. This is in contrast to the buccal cavity of the pharynx, that grows only during the ecdysis. Chitin, the stiff material that makes the buccal cavity hard, was also found in the lumen of the whole pharynx, suggesting a similar structural function in the pharynx, although the chitin was localized in more restricted regions of the lumen compared to chitinous structures such as the buccal cavity or the grinder (Zhang et al. 2005). Our observations on the continuous growth of the whole pharynx suggest that the lumen is still relatively flexible, despite the presence of chitin. The growth in the intermolt instead of in the molt could be related to incapability of the pharynx to grow during the renewal and shedding of the pharynx lumen cuticle. Moreover, one explanation could be that the energy resources are limited during the molt, when animals cannot feed, and growth of the pharynx would thus take away resources from the active process of renewal and shedding of the cuticle of the pharynx, and other processes in the rest of the body (Monsalve and Frand 2012; Lazetic and Fay 2017). Overall, our observations link molting cycle gene expression oscillations with growth of the pharynx. It would be interesting to find out if other organs in *C. elegans*, but also in other animals, are linked to dynamics in gene expression.

5.5 Material and methods

C. elegans strains and handling

Wild-type nematodes were strain N2. Mutants: *mgIs49[mlt-10p::gfp-pest; ttx-3p::gfp]* and *uos5[daf-16::mScarlet]* (starvation time-lapse experiments).

All strains were handled according to the standard protocol (Brenner 1974). Briefly, animals were grown on agar plates, consisting out of nematode growth medium (NGM), seeded with the *E. coli* strain OP50 as food source and grown at 20 °C.

Starvation was induced by washing the animals from plates with M9 buffer and transferred into a 15 mL tube. After that, the animals were washed 3x with M9 buffer for 5 minutes. The timer of starvation was started at the first wash step. After washing the bacteria of the animals, the animals were kept in the tube with M9 buffer, gently rocking until the starvation time was reached.

Single molecule fluorescence in situ hybridization

The oligonucleotides for the smFISH probe sets were designed with optimal GC content and specificity for the gene of interest using the Stellaris RNA FISH probe designer. The oligonucleotides were synthesized with a 3' amino C7 modification and purified by LGC Biosearch Technologies. Conjugation of the oligonucleotides with either Cy5 (GE Amersham) or Alexa594 (Invitrogen) was done as previously described (Lyubimova et al. 2013). The smFISH protocol was performed as previously described (Raj et al. 2008; Ji and van Oudenaarden 2012). Briefly, staged animals were washed from plates with M9 buffer and fixed in 4% formaldehyde in 1x PBS, gently rocking at room temperature (RT) for 40 minutes (young adults for 35 minutes). Fixation of embryos required a snap-freeze step to crack the eggshells by submerging embryos, after 15 minutes in fixation buffer, in liquid nitrogen, and thawing on ice for 20 minutes. After fixation, the animals were 2x washed with 1xPBS and resuspended in 70% ethanol overnight at 4°C. Ethanol was removed and animals were washed with 10% formamide and 2X SSC, as preparation for the hybridization. Animals were incubated with the smFISH probes overnight in the dark at 37°C in a hybridization solution (Stellaris) with added 10% formamide. The next day, animals were washed 2x with 10% formamide and 2X SSC each with an incubation step of 30 minutes at 37°C. The last wash step contains DAPI 5 µg/mL for nuclear staining. The wash buffer was removed, and animals were resuspended in 2X SSC and stored at 4°C until imaging. The 2X SSC was aspirated and animals were immersed in 100 µl GLOX buffer (0.4% glucose, 10 mM Tris-HCl, pH 8.0, 2X SSC) together with 1 µl Catalase (Sigma-Aldrich) and 1 µl glucose oxidase (Sigma-Aldrich) (3.7 mg/mL) to prevent bleaching during imaging.

Microscopy images of smFISH samples were acquired with a Nikon Ti-E inverted fluorescence microscope, equipped with a 100X plan-apochromat oil-immersion objective and an Andor Ikon-M CCD camera controlled by µManager software (Edelstein et al. 2014). smFISH analysis was performed with custom Python

software, based on a previously described method (Raj et al. 2008). Briefly, we first convolved the smFISH images with a Gaussian filter. Next, candidate spots were selected via manual thresholding, and partially overlapping spots were separated via finding 3D regional intensity maxima.

Time-lapse microscopy

Time-lapse imaging was performed as previously described (Gritti et al. 2016). Briefly, micro chambers are made out of polyacrylamide hydrogel, made from a 10% dilution of 29:1 acrylamide/bis-acrylamide (Sigma-Aldrich) was mixed with 0.1% ammonium persulfate (Sigma-Aldrich) and 0.01% TEMED (Sigma-Aldrich) as polymerization initiators. For the time-lapse experiments the chambers were 240x240x20 mm, these dimensions were sufficient to contain enough OP50 bacteria to sustain development until animals started laying eggs.

We used a Nikon Ti-E inverted microscope with a 40X objective in all experiments. The microscope has a Hamamatsu Orca Flash 4.0 v2 camera set at full frame and full speed. The camera chip is 13x13 mm and contains 4Mp. To keep track of the molting cycle as indication of the age of the animals and the growth of the pharynx, we used bright field imaging, which contained a red LED (CoolLED, 630nm). Time-lapse images were acquired every 20 minutes without detectable phototoxicity effects.

Starvation induction was achieved by placing less food in each of the chambers than required for normal development. Subsequently, the animals run out of food before finishing development. Bacteria are visible with the bright field imaging, but this could not provide enough information on when the animals induced their starvation response.

5.5.5 Analysis & statistics

All smFISH data was analyzed with custom written software in Python, as described in Chapter 2.

The RNA-sequencing data (specifically TimeCourse 2) from the study of Hendriks *et al.* (Hendriks et al. 2014) and the RNA-sequencing data from the study of Werner *et al.* (Werner et al. 2020) were processed and analyzed using R and python. Raw RNA-sequencing data from the study of Werner *et al.*, was processed as described, where 34 samples containing at least 10^5 and less than $3 \cdot 10^6$ total transcripts were used for follow-up analysis.

For the analysis of the Hendriks et al. data, we first made a subset of genes expressed in the pharynx, based on annotated single-cell RNA-sequencing data (Cao

et al. 2017). For the locating the molt and intermolt periods in the data from the study of Meeuse *et al.* (Meeuse et al. 2020), we used a list of APPG genes (WBGene00004099, WBGene00000030, WBGene00000029, WBGene00000031, WBGene00000024, WBGene00004139, WBGene00000034, WBGene00004141, WBGene00004170, WBGene00000028, WBGene00004171, WBGene00004158, WBGene00004174), to create a subset of genes that promote the assembly and function of the pharyngeal cuticle (George-Raizen et al. 2014), to guess the phase range of the timing of ecdysis. And we used a list of genes to create a subset of genes involved in renewal and shedding of the cuticle (WBGene00001064, WBGene00001065, WBGene00001069, WBGene00001070, WBGene00001072, WBGene00001067, WBGene00001074, WBGene00005016, WBGene00001076, WBGene00003057, WBGene00004397, WBGene00000251, WBGene00001071, WBGene00005017, WBGene00004398, WBGene00000252, WBGene00015646, WBGene00021095, WBGene00008605, WBGene00012186) (Lazetic and Fay 2017), to guess the phase range of the molt.

The measurements on body and pharynx size in the time-lapse microscopy images were performed using imageJ (Schneider, Rasband, and Eliceiri 2012).

For the GO term analysis on the groups of genes by the RNA-sequencing data from the study of Meeuse *et al.*, was conducted using Panther (Thomas et al. 2003). A statistical overrepresentation test was performed on each of the groups, with the standard reference list of the whole genome of *C. elegans*.

5.6 Appendix

5.6.1 Author contributions to the chapter

JJHT, TL and JSvZ conceptualized the study. JJHT and JSvZ designed the experimental strategy. All authors discussed and interpreted results. JJHT performed the experiments and analyzed the data. JJHT wrote the text. JSvZ reviewed and edited the text.

5.6.2 Supplementary figures

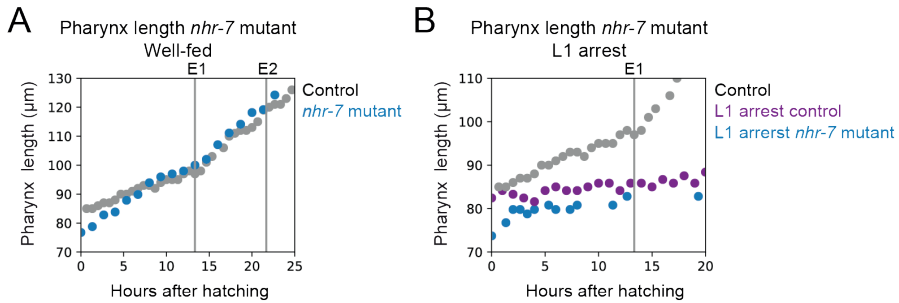


Figure S1: *nhr-7* mutants show normal pharynx growth and arrest. (A) Null-mutants of *nhr-7* show normal pharynx development on food compared to control animals in time-lapse setup with micro chambers (Gritti et al. 2016). (B) Null-mutants of *nhr-7* show a normal arrest of the pharynx growth upon hatching without food, and subsequently going into L1 arrest. Ecdysis in (A-B) are the ecdysis of the control animal.

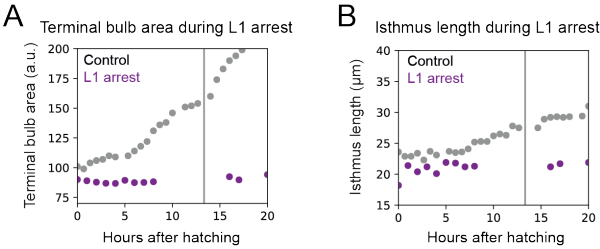


Figure S2: L1 arrested animals maintain similar proportions between pharynx regions during starvation. (A-B) The size of the terminal bulb and the width of the isthmus remain constant throughout the time that animals are in L1 arrest (0-20 hours).

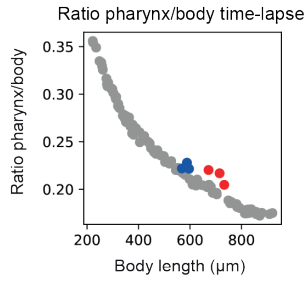


Figure S3: Larger pharynx in arrested animals in time-lapse data. Most animals had a slightly larger pharynx compared to the body at the end of the time-lapse experiment. Blue: L3 arrested animals. Red: L4/ARD arrested animals.

5.7 References

- Angelo, G., and M. R. Van Gilst. 2009. 'Starvation protects germline stem cells and extends reproductive longevity in *C. elegans*,' *Science*, 326: 954-8.
- Ardizzi, J. P., and H. F. Epstein. 1987. 'Immunochemical localization of myosin heavy chain isoforms and paramyosin in developmentally and structurally diverse muscle cell types of the nematode *Caenorhabditis elegans*,' *J Cell Biol*, 105: 2763-70.
- Baugh, L. R. 2013. 'To grow or not to grow: nutritional control of development during *Caenorhabditis elegans* L1 arrest,' *Genetics*, 194: 539-55.
- Brenner, S. 1974. 'The genetics of *Caenorhabditis elegans*,' *Genetics*, 77: 71-94.
- Cao, J., J. S. Packer, V. Ramani, D. A. Cusanovich, C. Huynh, R. Daza, X. Qiu, C. Lee, S. N. Furlan, F. J. Steemers, A. Adey, R. H. Waterston, C. Trapnell, and J. Shendure. 2017. 'Comprehensive single-cell transcriptional profiling of a multicellular organism,' *Science*, 357: 661-67.
- Chen, Jianjun, and Edward Caswell-Chen. 2003. 'Why *Caenorhabditis elegans* adults sacrifice their bodies to progeny,' *Nematology*, 5: 641.
- Dequeant, M. L., E. Glynn, K. Gaudenz, M. Wahl, J. Chen, A. Mushegian, and O. Pourquie. 2006. 'A complex oscillating network of signaling genes underlies the mouse segmentation clock,' *Science*, 314: 1595-8.
- Ebisuya, M., and J. Briscoe. 2018. 'What does time mean in development?,' *Development*, 145.
- Edelstein, A. D., M. A. Tsuchida, N. Amodaj, H. Pinkard, R. D. Vale, and N. Stuurman. 2014. 'Advanced methods of microscope control using muManager software,' *J Biol Methods*, 1.
- Frand, A. R., S. Russel, and G. Ruvkun. 2005. 'Functional genomic analysis of *C. elegans* molting,' *PLoS Biol*, 3: e312.
- George-Raizen, J. B., K. R. Shockley, N. F. Trojanowski, A. L. Lamb, and D. M. Raizen. 2014. 'Dynamically-expressed prion-like proteins form a cuticle in the pharynx of *Caenorhabditis elegans*,' *Biol Open*, 3: 1139-49.
- Gritti, N., S. Kienle, O. Filina, and J. S. van Zon. 2016. 'Long-term time-lapse microscopy of *C. elegans* post-embryonic development,' *Nat Commun*, 7: 12500.
- Grun, D., M. Kirchner, N. Thierfelder, M. Stoeckius, M. Selbach, and N. Rajewsky. 2014. 'Conservation of mRNA and protein expression during development of *C. elegans*,' *Cell Rep*, 6: 565-77.
- Harvald, E. B., R. R. Sprenger, K. B. Dall, C. S. Ejsing, R. Nielsen, S. Mandrup, A. B. Murillo, M. Larance, A. Gartner, A. I. Lamond, and N. J. Faergeman. 2017. 'Multi-omics Analyses of Starvation Responses Reveal a Central Role for Lipoprotein Metabolism in Acute Starvation Survival in *C. elegans*,' *Cell Syst*, 5: 38-52 e4.
- Hendriks, G. J., D. Gaidatzis, F. Aeschmann, and H. Grosshans. 2014. 'Extensive oscillatory gene expression during *C. elegans* larval development,' *Mol Cell*, 53: 380-92.
- Ji, N., and A. van Oudenaarden. 2012. 'Single molecule fluorescent in situ hybridization (smFISH) of *C. elegans* worms and embryos,' *WormBook*: 1-16.
- Kageyama, R., Y. Niwa, A. Isomura, A. Gonzalez, and Y. Harima. 2012. 'Oscillatory gene expression and somitogenesis,' *Wiley Interdiscip Rev Dev Biol*, 1: 629-41.
- Kamal, M., H. Moshiri, L. Magomedova, D. Han, K. C. Q. Nguyen, M. Yeo, J. Knox, R. Bagg, A. M. Won, K. Szlapa, C. M. Yip, C. L. Cummins, D. H. Hall, and P. J. Roy. 2019. 'The marginal cells of the *Caenorhabditis elegans* pharynx scavenge cholesterol and other hydrophobic small molecules,' *Nat Commun*, 10: 3938.
- Keil, W., L. M. Kutscher, S. Shaham, and E. D. Siggia. 2017. 'Long-Term High-Resolution Imaging of Developing *C. elegans* Larvae with Microfluidics,' *Dev Cell*, 40: 202-14.

- Killian, D. J., and E. J. Hubbard. 2005. 'Caenorhabditis elegans germline patterning requires coordinated development of the somatic gonadal sheath and the germ line', *Dev Biol*, 279: 322-35.
- Kim, Dh, D. Grun, and A. van Oudenaarden. 2013. 'Dampening of expression oscillations by synchronous regulation of a microRNA and its target', *Nat Genet*, 45: 1337-44.
- Kimble, J. E., and J. G. White. 1981. 'On the control of germ cell development in *Caenorhabditis elegans*', *Dev Biol*, 81: 208-19.
- Knight, C. G., M. N. Patel, R. B. Azevedo, and A. M. Leroi. 2002. 'A novel mode of ecdysozoan growth in *Caenorhabditis elegans*', *Evol Dev*, 4: 16-27.
- Kostal, V. 2006. 'Eco-physiological phases of insect diapause', *J Insect Physiol*, 52: 113-27.
- Krol, A. J., D. Roellig, M. L. Dequeant, O. Tassy, E. Glynn, G. Hattem, A. Mushegian, A. C. Oates, and O. Pourquie. 2011. 'Evolutionary plasticity of segmentation clock networks', *Development*, 138: 2783-92.
- Kruse, K., and F. Julicher. 2005. 'Oscillations in cell biology', *Curr Opin Cell Biol*, 17: 20-6.
- Lazetic, V., and D. S. Fay. 2017. 'Molting in *C. elegans*', *Worm*, 6: e1330246.
- Lewis, J. A., and J. T. Fleming. 1995. 'Basic culture methods', *Methods Cell Biol*, 48: 3-29.
- Lyubimova, A., S. Itzkovitz, J. P. Junker, Z. P. Fan, X. Wu, and A. van Oudenaarden. 2013. 'Single-molecule mRNA detection and counting in mammalian tissue', *Nat Protoc*, 8: 1743-58.
- Matsuda, M., Y. Yamanaka, M. Uemura, M. Osawa, M. K. Saito, A. Nagahashi, M. Nishio, L. Guo, S. Ikegawa, S. Sakurai, S. Kihara, T. L. Maurissen, M. Nakamura, T. Matsumoto, H. Yoshitomi, M. Ikeya, N. Kawakami, T. Yamamoto, K. Woltjen, M. Ebisuya, J. Toguchida, and C. Alev. 2020. 'Recapitulating the human segmentation clock with pluripotent stem cells', *Nature*, 580: 124-29.
- Meeuse, M. W., Y. P. Hauser, L. J. Morales Moya, G. J. Hendriks, J. Eglinger, G. Bogaarts, C. Tsiarlis, and H. Grosshans. 2020. 'Developmental function and state transitions of a gene expression oscillator in *Caenorhabditis elegans*', *Mol Syst Biol*, 16: e9498.
- Meli, V. S., B. Osuna, G. Ruvkun, and A. R. Frand. 2010. 'MLT-10 defines a family of DUF644 and proline-rich repeat proteins involved in the molting cycle of *Caenorhabditis elegans*', *Mol Biol Cell*, 21: 1648-61.
- Miller, D. M., F. E. Stockdale, and J. Karn. 1986. 'Immunological identification of the genes encoding the four myosin heavy chain isoforms of *Caenorhabditis elegans*', *Proc Natl Acad Sci U S A*, 83: 2305-9.
- Mok, D. Z., P. W. Sternberg, and T. Inoue. 2015. 'Morphologically defined sub-stages of *C. elegans* vulval development in the fourth larval stage', *BMC Dev Biol*, 15: 26.
- Monsalve, G. C., and A. R. Frand. 2012. 'Toward a unified model of developmental timing: A "molting" approach', *Worm*, 1: 221-30.
- Morris, C. A., L. D. Morris, A. R. Kennedy, and H. L. Sweeney. 2005. 'Attenuation of skeletal muscle atrophy via protease inhibition', *J Appl Physiol (1985)*, 99: 1719-27.
- Raizen, D. M., J. E. Zimmerman, M. H. Maycock, U. D. Ta, Y. J. You, M. V. Sundaram, and A. I. Pack. 2008. 'Lethargus is a *Caenorhabditis elegans* sleep-like state', *Nature*, 451: 569-72.
- Raj, A., P. van den Bogaard, S. A. Rifkin, A. van Oudenaarden, and S. Tyagi. 2008. 'Imaging individual mRNA molecules using multiple singly labeled probes', *Nat Methods*, 5: 877-9.
- Saini, R., M. Jaskolski, and S. J. Davis. 2019. 'Circadian oscillator proteins across the kingdoms of life: structural aspects', *BMC Biol*, 17: 13.
- Schindler, A. J., L. R. Baugh, and D. R. Sherwood. 2014. 'Identification of late larval stage developmental checkpoints in *Caenorhabditis elegans* regulated by insulin/IGF and steroid hormone signaling pathways', *PLoS Genet*, 10: e1004426.

- Schneider, C. A., W. S. Rasband, and K. W. Eliceiri. 2012. 'NIH Image to ImageJ: 25 years of image analysis,' *Nat Methods*, 9: 671-5.
- Schulenburg, H., and M. A. Felix. 2017. 'The Natural Biotic Environment of *Caenorhabditis elegans*,' *Genetics*, 206: 55-86.
- Singh, R.N., and J.E. Sulston. 1978. 'Some Observations On Moulting in *Caenorhabditis Elegans*,' *Nematologica*, 24: 63.
- Thomas, P. D., M. J. Campbell, A. Kejariwal, H. Mi, B. Karlak, R. Daverman, K. Diemer, A. Muruganujan, and A. Narechania. 2003. 'PANTHER: a library of protein families and subfamilies indexed by function,' *Genome Res*, 13: 2129-41.
- Turek, M., and H. Bringmann. 2014. 'Gene expression changes of *Caenorhabditis elegans* larvae during molting and sleep-like lethargus,' *PLoS One*, 9: e113269.
- Uriu, K. 2016. 'Genetic oscillators in development,' *Dev Growth Differ*, 58: 16-30.
- Werner, Steffen, W Mathijs Rozemuller, Annabel Ebbing, Anna Alemany, Joleen Traets, Jeroen S. van Zon, Alexander van Oudenaarden, Hendrik C. Korswagen, Greg J. Stephens, and Thomas S. Shimizu. 2020. 'Functional modules from variable genes: Leveraging percolation to analyze noisy, high-dimensional data,' *bioRxiv*: 2020.06.10.143743.
- Zhang, Y., J. M. Foster, L. S. Nelson, D. Ma, and C. K. Carlow. 2005. 'The chitin synthase genes *chs-1* and *chs-2* are essential for *C. elegans* development and responsible for chitin deposition in the eggshell and pharynx, respectively,' *Dev Biol*, 285: 330-9.

Chapter 6

General discussion & outlook

In this thesis, we studied two key challenges that cells face throughout their entire lifetime: first, the timing of the developmental events involved, and when the cell reached the terminal differentiated, this identity needs to be maintained. We used *Caenorhabditis elegans* as a model system to study the molecular mechanism by which both challenges are addressed during animal development.

The work in this thesis relied extensively on quantifying gene expression with high precision. For this, we used single molecule FISH (smFISH), a technique to visualize individual mRNA molecules. In **Chapter 2**, I described the WormFISH software package, that we developed to optimize the workflow of analyzing large smFISH data sets on *C. elegans* larvae. In addition, we developed tools for the package to annotate the developmental stage of individual animals, which allows for using smFISH data to study expression dynamics in time.

Next, we studied the maintenance of terminal cell fate in neuronal cells that are relatively long-lived, since they are hardly renewed during the lifetime of an animal. In **Chapter 3**, we identified a mechanism for lifelong maintenance of stable cell identity and function in a neuron pair involved in the chemotaxis to salt in *C. elegans*.

In **Chapter 4**, we attempted to validate the existence and the function of two clusters, one containing neuron-specific and the other pharynx-specific genes, that we found with our novel percolation clustering approach in single-animal RNA sequencing data. We quantified the expression levels of pairs of selected genes from both clusters, by which we validated the correlated variability of one of the clusters. Moreover, we also found a possible explanation of the correlation between the clusters related to a developmental stage dependent starvation response.

The challenge of time-keeping of developmental events is addressed in **Chapter 5**. There we found that the oscillatory gene expression during development, which has previously been thought to be mainly involved in the process of molting, is also functionally linked with the growth of the pharynx. This result suggests that the oscillator, which is coupled to the molting cycle, is also involved in developmental processes other than the molt.

In this general discussion, we will summarize and discuss our most important findings, together with the outlook. Moreover, we will also discuss the potential impact and implications for society.

6.1 Long-term maintenance of ASE fate with the CHE-1 switch

Terminal selector genes are transcription factors that are responsible for the terminal cell fate by inducing a battery of genes giving rise to the terminal cell fate identity. Crucially, those terminal selector genes also induce their own expression via autoregulation, making them act as genetic cell fate switches. While some cells are renewed every few days, neuronal cells are hardly renewed and have to maintain their terminal cell identity for the entire lifetime of an organism (Ming and Song 2005; Spalding et al. 2005). How the terminal cell identity is maintained over such long timescales despite intrinsic molecular fluctuations in the components of the genetic cell fate switch, is not known. In **Chapter 3**, we focused on the ASE neurons in the nematode *C. elegans*, in which the terminal selector gene *che-1* induces both its own expression and that of hundreds of ASE subtype specific target genes. We identified a novel mechanism that explains life-long maintenance of ASE neuron fate by the terminal selector transcription factor *che-1*, called *targeted-reservoir buffering*. The fluctuations of CHE-1 levels are buffered by the reservoir of CHE-1 bound at its target promoters, by preferential binding of CHE-1 to its own promoter. This mechanism ensures the continued expression of *che-1* over severely long time-scales. We could experimentally validate this mechanism by transiently depleting the CHE-1 levels. During transient depletion of CHE-1 levels, the *che-1* expression was resilient, while the target genes of CHE-1 together with the ASE function were lost. Additionally, we discovered a 130 bp fragment in the *che-1* promoter region responsible for the stable *che-1* expression. Moreover, we also identified an *Otx*-related homeodomain binding site, which caused stochastic loss of ASE identity long after determination in the embryonic stage.

So far, we do not have direct experimental evidence of the *targeted-reservoir buffering* mechanism. ChIP-sequencing experiments have proven to be difficult due to only ~900 CHE-1 proteins in each of the two ASE neurons. In addition, we do not know what the *Otx*-related TF or co-factors involved are, making *in vitro* measurements on the kinetics or affinity related to the preferential binding of CHE-1 on its own promoter impossible. A recent study from Gurdon *et al.* (Gurdon et al. 2020), identified another similar mechanism by which the two lineage-determining factors *MyoD* and *Ascl1* (similar as terminal selectors) maintain long-term expression. Instead of fast binding and unbinding rates from the DNA binding site, both *MyoD* and *Ascl1* can remain bound for several hours or even days. These long dwell times could be used as a strategy to maintain differentiated states. The possibility of similarly long binding times of CHE-1 cannot be rejected by our current experimental data. The conclusions are similar to our findings with the mathematical model, in which preference to the CHE-1 promoter can also explain the long-term maintenance of cell identity. A dissimilarity is the extensive binding time of *MyoD* and

Ascl1 of days, while in our mathematical model 100 sec on the *che-1* promoter versus 1 sec on the target gene promoters is already sufficient. However, the experiments in this study on *MyoD* and *Ascl1* were performed in *Xenopus* oocytes, and not in differentiated cells of an organism, leaving open what the kinetics are in differentiated cells. It would be interesting to directly measure the *in vivo* dwell time of CHE-1 within the ASE neurons. However, measuring *in vivo* dwell times would require experiments, such as *in vivo* single-molecule tracking, that might prove to be hard due to autofluorescence and tissue scattering in the ASE neurons of *C. elegans*.

We showed that the stability of the maintenance of the *che-1* expression depends on the *Otx*-related binding motif close to the ASE motif in the *che-1* promoter. The *Otx*-related binding motif is similar to the motif on which CEH-36 is known to bind. However, the *ceh-36* loss-of-function mutant did not show the same phenotype observed in the *(DHD)p::che-1* animals, with a deleted *Otx*-related binding motif in the *che-1* promoter. Homeodomain proteins are found in a wide range of species, of which several are known to be involved in specifying the identity of neuronal subtypes (Germot et al. 2001; Ang et al. 1996; Simeone, Puellas, and Acampora 2002). A recent study of Reilly et al. (Reilly et al. 2020), showed that the diversity of neuronal subtypes in *C. elegans* could be grouped by unique combinations of homeodomain proteins. Their analysis on several genetic loss-of-function homeodomain mutants indicated that many homeodomain proteins could also have a critical role in specifying neuronal identity. Homeodomain proteins have been shown before to function as terminal selector genes of neuronal subtypes in *C. elegans*, but also as proteins that stabilize the expression of the terminal selector gene. The stabilizing function of homeodomain proteins has been described in the TRN (touch receptor neurons), in which the expression of *mec-3* is stabilized with aid of the homeodomain protein ALR-1 by restricting variability of *mec-3* expression (Topalidou, van Oudenaarden, and Chalfie 2011). These observations suggest that a mechanism, in which additional homeodomain proteins aid in stabilizing terminal cell fate, could be a common shared strategy between neuronal subtypes.

We do not know if and which homeodomain protein binds the *Otx*-related binding motif on the *che-1* promoter. We already checked a number of mutants of homeodomain proteins present in ASE neurons (Cao et al. 2017), such as *ceh-23*, *ceh-36*, *alr-1* and *dsc-1* (data not shown). None of these mutants showed similar phenotypes as the *Otx*-related binding motif deletion mutant. This either suggests that other homeodomain proteins are involved or that the function of the homeodomain proteins is redundant. A mutant screen on all of the nine homeodomain proteins expressed in the ASE neurons, including single, double, potentially up to nonuplet mutants, could reveal which homeodomain protein(s) are responsible. However, potentially a non-homeodomain protein could bind as well. To find out

if other proteins with the potential of binding on the specific DNA-sequence are involved, isolation of bound proteins to the *che-1* promoter could identify potential candidates. This could be achieved by proteomics of isolated chromatin segments (PIch) or by chromatin immunoprecipitation with mass-spectrometry (ChIP-MS), by isolation of the *che-1* promoter region together with the bound transcription factors. When we have identified the transcription factors bound on the *che-1* promoter, this would also allow for studying the binding kinetics and affinities (e.g. with the ElectroMobility Shift Assay, EMSA), to study how the *Otx*-related binding motif on the *che-1* promoter aids in stabilizing *che-1* expression.

When CHE-1 proteins are depleted sufficiently long enough, the markers for the ASE fate and the ASE function are lost, even when the depletion is stopped. However, we do not know what identity the cells have after loss of *che-1* expression. It might revert back to the cell identity of a neuronal precursor cell, or it could potentially obtain the terminal cell fate of another neuronal subtype. The latter would require cellular plasticity of the former ASE neurons. Moreover, generally, cellular plasticity is restricted by chromatin remodeling to prevent other terminal selectors from inducing their target genes in the differentiated neuronal cells (Patel and Hobert 2017), suggesting that the chromatin state would restrict the potential of the cells to adapt another terminal cell fate. However, it has been suggested that the transcription factor *die-1* in the ASE neurons operates together with the MYST complex to initiate and maintain left versus right ASE subtype (O'Meara, Zhang, and Hobert 2010). Moreover, with temperature sensitive mutants, the MYST complex was shown to maintain the left ASE fate. This suggests that the chromatin state of those cell-specific genes needs continuous presence of recruiting transcription factors. The loss of *che-1* expression could potentially lead to changes of specific chromatin states if the expression of the target gene *die-1* is also lost, but it is unknown if *die-1* is lost and what the effect would be on the cellular plasticity.

It would be interesting to test if the mechanism of *targeted-reservoir buffering* for maintaining the terminal cell fate also works in more complex networks that rely on similar terminal selector genes. For example, in the AIY interneurons, involved in processing sensory information and in behavioral plasticity (Tsalik and Hobert 2003; Remy and Hobert 2005), two terminal selector genes are responsible for the AIY cell fate: the TTX-3 LIM homeodomain protein and the CEH-10 Paired-type homeodomain protein (Wenick and Hobert 2004). Similar to the CHE-1 target genes, the AIY cell-specific genes all share the same AIY binding motif in their promoter region. This AIY motif is synergistically activated by a dimer of the two homeodomain proteins. Crucially, the two terminal selector genes also induce their own expression, suggesting even more sensitivity to stochastic fluctuations compared to the simpler positive autoregulation of CHE-1. A similar approach as we applied to

the *che-1* switch, by measuring the levels and rates of the components, stochastic modelling and depletion experiments, could reveal the sensitivity of this more complex system for stochastic fluctuations and its strategy to cope with them.

We showed that the ASE fate is lost progressively over time in the *Otx*-related binding site mutants in the *che-1* promoter. The *Otx*-related binding site is necessary for stabilizing the expression levels of *che-1*. It has been suggested that mutations in the regulatory elements of DNA, such as the binding sites of TFs, are common in a diverse range of vertebrate diseases (Maurano et al. 2012; Douglas and Hill 2014). For example, the neuronal eye disease causing abnormal formation of the iris, Aniridia (MIM 106210), can be caused by a point mutation in the regulatory element essential in maintaining the expression of *PAX6*, a Paired (Pax) family gene essential in the eye, brain, olfactory and pancreas development (Hill et al. 1991; Stoykova et al. 1996; St-Onge et al. 1997; Axton et al. 1997; Lee, Khan, and O’Keefe 2008). Moreover, a particular point mutation in the *PAX6* binding site has been shown to disrupt the positive autoregulation of *PAX6* during the development of the eye, causing Aniridia (Bhatia et al. 2013). Neuronal disease can also be progressive and lead to loss of function of tissue later on in life. Progressive neuronal diseases, such as cone dysfunction syndromes causing progressive abnormal vision or even loss of sight (Michaelides, Hunt, and Moore 2004), most likely does not involve mutations disrupting the functioning of the terminal selector gene instantly, since this would lead to immediate loss of function of the tissue after development. Instead, mutations affecting the autoregulation and the functioning of additional factors that stabilize the expression of the terminal selector gene could potentially contribute to progressive neuronal diseases. A better understanding on how additional factors stabilize the expression of the terminal selectors could potentially make a valuable contribution to our understanding of the mechanistic variety of the causes of progressive neuronal diseases.

6.2 Growth of the pharynx coordinated with molting cycle oscillator

In **Chapter 4**, we validated the existence of one of the clusters found with the novel percolation clustering approach on single-animal RNA-sequencing data. Moreover, this cluster consists out of genes that are specifically expressed in the pharynx and those genes show oscillatory behavior in gene expression related to development. In addition, we found with gene analysis that many genes in the cluster are structural components or related to the functioning of the pharynx muscles. Previously, the wide-spread oscillatory gene expression during *C. elegans* development has been thought to be mainly involved in the molting cycle (Hendriks et al. 2014; Meeuse et al. 2020). Interestingly, most of the genes in the cluster had a peak phase outside the molting period. These observations attracted our interest to follow the cluster up,

and to study the role of oscillatory gene expression in the pharynx.

In **Chapter 5**, we showed that a group of genes, including most genes from the cluster, expressed in the pharynx with oscillatory behavior that have a phase peak outside the molting period, appear to be functionally linked with the growth of the pharynx. Moreover, we followed the growth of the pharynx in time and showed that the growth of the pharynx has a cyclical character. The oscillating expression of the selected genes in the pharynx appeared to be coordinated with the growth dynamics of the pharynx in development. When we arrested the development at different stages through immediate starvation, the pharynx growth halted and similarly also the oscillations in gene expression stopped, suggesting that both are functionally coupled.

Initially, the oscillating gene expression in *C. elegans* was thought to be only related to the molting cycle. Yet, many genes with oscillating gene expression were found with a phase outside the molt (Meeuse et al. 2020). It has been suggested that those phases could also be a byproduct without clear function, and occur as a side effect of the periodic molting process (Unpublished: Tsiaris & Großhans 2020). However, we have shown that genes with a specific phase outside the molt are coordinated with the growth of the pharynx, and not directly related to the molt. Apart from a direct link with development, we also found oscillating genes in the pharynx that were more indirectly related to development, such as genes related to defense against bacteria and muscle activity. In a previous study of Turek *et al.* (Turek and Bringmann 2014), they suggest that the oscillating gene expression could also be involved in the cycle of active- (intermot) and inactive-stages (lethargus) during development of the animals. Circadian clocks in animals have been shown to anticipate in food-cycles, except those rely on the day and night rhythm (Patke, Young, and Axelrod 2020). The oscillating gene expression in *C. elegans* could also fulfill a similar role as a food-cycle, without influence of external time, but instead as a function of developmental processes. However, if the oscillator determines or only aids in the all previously mentioned developmental processes, remains unclear.

So far, we do not know what regulates the oscillations in gene expression in the pharynx. At least three pharyngeal myogenic transcription factors, *peb-1*, *ceh-22* and *pha-4* are known to control the expression of the pharyngeal myosin heavy chains, *myo-1* and *myo-2* (Kalb et al. 2002; Gaudet and Mango 2002; Okkema et al. 1997), and potentially also the expression of *nhr-7* and *marg-1* based on bioinformatics analysis. Both *ceh-22* and *pha-4* do not oscillating during development, while *peb-1* does oscillate during development. Moreover, it has been suggested that PEB-1 negatively modulates the ability of PHA-4 to activate transcription (Kalb et al. 2002). Interestingly, *peb-1* oscillates during development in a peak phase right

after the peak phase of the growth-related genes, including *myo-1* and *myo-2* (data not shown). Subsequently, the oscillating expression of *peb-1* and its negative regulation on the expression of *myo-1* and *myo-2*, could potentially cause the observed oscillations in gene expression of those genes, and potentially other growth-related genes in the pharynx. So far, it is known that mutants of *peb-1* are particularly defective in shedding of pharyngeal cuticle, and most animals go into developmental arrest (Fernandez, Gibbons, and Okkema 2004). It would be interesting to measure the growth of the pharynx and quantify the expression levels of *myo-1* and *myo-2* in the *peb-1* mutants, to find out if the cyclical character of growth of the pharynx and the oscillations in gene expression of *myo-1* and *myo-2* are also affected.

We found that the pharynx size retains stereotypical dynamics in the ratio with the body size in *C. elegans* throughout development. The scaling relationship between the whole body and body parts is also called allometry, in which the allometric traits, such as the heart or the brain, can grow at different rates than the body (Bonner 2011). Growth of organs and the body as a whole, is achieved by cell divisions and increase of cell volume. Over the past decades, studies have shown that animals can regulate the overall size of their tissues and organs by monitoring their absolute size by the cell numbers and cell volume (Day and Lawrence 2000; Vervoort et al. 1999; Fankhauser 1945; Potter and Xu 2001). In *C. elegans*, the postnatal growth is mainly driven by increase of cell volume instead of by cell division. The global body size is regulated by the TGF- β /Sma/Mab pathway, in which the pathway is activated by a ligand in the TGF- β family, called DBL-1 (Tuck 2014; Dineen and Gaudet 2014). The SMAD family are the downstream effectors in this pathway. When *sma-3*, a member of the SMAD family which is expressed in the pharynx, is mutated, the body and pharynx size are smaller. Interestingly, they found that when *sma-3* mutants were rescued solely by expression of the SMA-3 in the pharynx, both pharynx and body size became wildtype (Dineen and Gaudet 2014). These observations suggest a strong coupling between the pharynx and body size. In contrast, we found decoupling of the pharynx and body growth when the animals are severely starved causing arrest of development. Not much is known in *C. elegans* about the regulation of growth and maintaining proportionality of organs, such as the pharynx. The decoupling during starvation suggests that there is no regulation mechanism of maintained allometry of the pharynx upon starvation. One explanation could be that there are no negative effects of having a larger pharynx during arrest, and halting growth of the rest of the body apart from the pharynx leaves enough energy resources to reach the developmental checkpoint. It would be interesting to show if erroneous continuation of development during starvation beyond developmental checkpoints, such as in *daf-16* mutants (Schindler, Baugh, and Sherwood 2014), causes the pharynx also to continue its growth. In addition, it remains unclear how and if the growth of the pharynx is controlled and monitored in relation to the

growth of the whole body. Therefore, it would be interesting to halt the growth of the pharynx without directly affecting growth of the rest of the body, while monitoring the body and pharynx size.

In conclusion, I discussed the two key findings in this thesis on timing of developmental events and maintenance of terminal cell identity. In the first part of the discussion, we focused on the life-long maintenance of terminal cell identity of the ASE neurons by the *che-1* switch. We found a novel mechanism in which preferential binding of *che-1* to its own promoter ensures the continued expression of *che-1* over severely long time-scales, which we could experimentally validate by depletion experiments. We also showed that a co-factor is involved in stabilizing the *che-1* expression. We speculated that non-coding mutations, related to interrupting stability of terminal selector by co-factors, could potentially lead to progressive neuronal diseases. Future studies on how co-factors stabilize the expression of terminal selectors, could potentially contribute to our understanding of the mechanistic variety of the causes of those progressive neuronal diseases. In the second part of the discussion, we focused on the functional link between growth of the pharynx and oscillations in gene expression expressed in the pharynx related to growth. Moreover, the pharynx has a different growth rate compared to the whole-body during development. However, this allometric relation is broken when animals are subjected to starvation. The control of allometry in the pharynx of *C. elegans* could provide us a relatively simple model system to study more general principles on the coupling between organ growth and the whole-body growth and how it this coupling is lost.

6.3 References

- Ang, S. L., O. Jin, M. Rhinn, N. Daigle, L. Stevenson, and J. Rossant. 1996. 'A targeted mouse *Otx2* mutation leads to severe defects in gastrulation and formation of axial mesoderm and to deletion of rostral brain,' *Development*, 122: 243-52.
- Axton, R., I. Hanson, S. Danes, G. Sellar, V. van Heyningen, and J. Prosser. 1997. 'The incidence of PAX6 mutation in patients with simple aniridia: an evaluation of mutation detection in 12 cases,' *J Med Genet*, 34: 279-86.
- Bhatia, S., H. Bengani, M. Fish, A. Brown, M. T. Divizia, R. de Marco, G. Damante, R. Grainger, V. van Heyningen, and D. A. Kleinjan. 2013. 'Disruption of autoregulatory feedback by a mutation in a remote, ultraconserved PAX6 enhancer causes aniridia,' *Am J Hum Genet*, 93: 1126-34.
- Bonner, John Tyler. 2011. *Why size matters: from bacteria to blue whales* (Princeton University Press).
- Cao, J., J. S. Packer, V. Ramani, D. A. Cusanovich, C. Huynh, R. Daza, X. Qiu, C. Lee, S. N. Furlan, F. J. Steemers, A. Adey, R. H. Waterston, C. Trapnell, and J. Shendure. 2017. 'Comprehensive single-cell transcriptional profiling of a multicellular organism,' *Science*, 357: 661-67.
- Day, S. J., and P. A. Lawrence. 2000. 'Measuring dimensions: the regulation of size and shape,' *Development*, 127: 2977-87.
- Dineen, A., and J. Gaudet. 2014. 'TGF-beta signaling can act from multiple tissues to regulate *C. elegans* body size,' *BMC Dev Biol*, 14: 43.
- Douglas, A. T., and R. D. Hill. 2014. 'Variation in vertebrate cis-regulatory elements in evolution and disease,' *Transcription*, 5: e28848.
- Fankhauser, G. 1945. 'Maintenance of normal structure in heteroploid salamander larvae, through compensation of changes in cell size by adjustment of cell number and cell shape,' *J Exp Zool*, 100: 445-55.
- Fernandez, A. P., J. Gibbons, and P. G. Okkema. 2004. '*C. elegans* *peb-1* mutants exhibit pleiotropic defects in molting, feeding, and morphology,' *Dev Biol*, 276: 352-66.
- Gaudet, J., and S. E. Mango. 2002. 'Regulation of organogenesis by the *Caenorhabditis elegans* FoxA protein PHA-4,' *Science*, 295: 821-5.
- Germot, A., G. Lecointre, J. L. Plouhinec, C. Le Mentec, F. Girardot, and S. Mazan. 2001. 'Structural evolution of *Otx* genes in craniates,' *Mol Biol Evol*, 18: 1668-78.
- Gurdon, J. B., K. Javed, M. Vodnala, and N. Garrett. 2020. 'Long-term association of a transcription factor with its chromatin binding site can stabilize gene expression and cell fate commitment,' *Proc Natl Acad Sci U S A*, 117: 15075-84.
- Hendriks, G. J., D. Gaidatzis, F. Aeschmann, and H. Grosshans. 2014. 'Extensive oscillatory gene expression during *C. elegans* larval development,' *Mol Cell*, 53: 380-92.
- Hill, R. E., J. Favor, B. L. Hogan, C. C. Ton, G. F. Saunders, I. M. Hanson, J. Prosser, T. Jordan, N. D. Hastie, and V. van Heyningen. 1991. 'Mouse small eye results from mutations in a paired-like homeobox-containing gene,' *Nature*, 354: 522-5.
- Kalb, J. M., L. Beaster-Jones, A. P. Fernandez, P. G. Okkema, B. Goszczynski, and J. D. McGhee. 2002. 'Interference between the PHA-4 and PEB-1 transcription factors in formation of the *Caenorhabditis elegans* pharynx,' *J Mol Biol*, 320: 697-704.
- Lee, H., R. Khan, and M. O'Keefe. 2008. 'Aniridia: current pathology and management,' *Acta Ophthalmol*, 86: 708-15.
- Maurano, M. T., R. Humbert, E. Rynes, R. E. Thurman, E. Haugen, H. Wang, A. P. Reynolds, R. Sandstrom, H. Qu, J. Brody, A. Shafer, F. Neri, K. Lee, T. Kutayavin, S. Stehling-Sun, A. K. Johnson, T. K. Canfield, E. Giste, M. Diegel, D. Bates, R. S. Hansen, S. Neph, P. J. Sabo, S. Heimfeld, A. Raubitschek, S. Ziegler, C. Cotsapas, N. Sotoodehnia, I. Glass,

- S. R. Sunyaev, R. Kaul, and J. A. Stamatoyannopoulos. 2012. 'Systematic localization of common disease-associated variation in regulatory DNA', *Science*, 337: 1190-5.
- Meeuse, M. W., Y. P. Hauser, L. J. Morales Moya, G. J. Hendriks, J. Eglinger, G. Bogaarts, C. Tsiairis, and H. Grosshans. 2020. 'Developmental function and state transitions of a gene expression oscillator in *Caenorhabditis elegans*', *Mol Syst Biol*, 16: e9498.
- Michaelides, M., D. M. Hunt, and A. T. Moore. 2004. 'The cone dysfunction syndromes', *Br J Ophthalmol*, 88: 291-7.
- Ming, G. L., and H. Song. 2005. 'Adult neurogenesis in the mammalian central nervous system', *Annu Rev Neurosci*, 28: 223-50.
- O'Meara, M. M., F. Zhang, and O. Hobert. 2010. 'Maintenance of neuronal laterality in *Caenorhabditis elegans* through MYST histone acetyltransferase complex components LSY-12, LSY-13 and LIN-49', *Genetics*, 186: 1497-502.
- Okkema, P. G., E. Ha, C. Haun, W. Chen, and A. Fire. 1997. 'The *Caenorhabditis elegans* NK-2 homeobox gene *ceh-22* activates pharyngeal muscle gene expression in combination with *pha-1* and is required for normal pharyngeal development', *Development*, 124: 3965-73.
- Patel, T., and O. Hobert. 2017. 'Coordinated control of terminal differentiation and restriction of cellular plasticity', *Elife*, 6.
- Patke, A., M. W. Young, and S. Axelrod. 2020. 'Molecular mechanisms and physiological importance of circadian rhythms', *Nat Rev Mol Cell Biol*, 21: 67-84.
- Potter, C. J., and T. Xu. 2001. 'Mechanisms of size control', *Curr Opin Genet Dev*, 11: 279-86.
- Reilly, M. B., C. Cros, E. Varol, E. Yemini, and O. Hobert. 2020. 'Unique homeobox codes delineate all the neuron classes of *C. elegans*', *Nature*, 584: 595-601.
- Remy, J. J., and O. Hobert. 2005. 'An interneuronal chemoreceptor required for olfactory imprinting in *C. elegans*', *Science*, 309: 787-90.
- Schindler, A. J., L. R. Baugh, and D. R. Sherwood. 2014. 'Identification of late larval stage developmental checkpoints in *Caenorhabditis elegans* regulated by insulin/IGF and steroid hormone signaling pathways', *PLoS Genet*, 10: e1004426.
- Simeone, A., E. Puelles, and D. Acampora. 2002. 'The *Otx* family', *Curr Opin Genet Dev*, 12: 409-15.
- Spalding, K. L., R. D. Bhardwaj, B. A. Buchholz, H. Druid, and J. Frisen. 2005. 'Retrospective birth dating of cells in humans', *Cell*, 122: 133-43.
- St-Onge, L., B. Sosa-Pineda, K. Chowdhury, A. Mansouri, and P. Gruss. 1997. 'Pax6 is required for differentiation of glucagon-producing alpha-cells in mouse pancreas', *Nature*, 387: 406-9.
- Stoykova, A., R. Fritsch, C. Walther, and P. Gruss. 1996. 'Forebrain patterning defects in Small eye mutant mice', *Development*, 122: 3453-65.
- Topalidou, I., A. van Oudenaarden, and M. Chalfie. 2011. 'Caenorhabditis elegans *aristless/Arx* gene *alr-1* restricts variable gene expression', *Proc Natl Acad Sci U S A*, 108: 4063-8.
- Tsalik, E. L., and O. Hobert. 2003. 'Functional mapping of neurons that control locomotory behavior in *Caenorhabditis elegans*', *J Neurobiol*, 56: 178-97.
- Tuck, S. 2014. 'The control of cell growth and body size in *Caenorhabditis elegans*', *Exp Cell Res*, 321: 71-6.
- Turek, M., and H. Bringmann. 2014. 'Gene expression changes of *Caenorhabditis elegans* larvae during molting and sleep-like lethargus', *PLoS One*, 9: e113269.
- Vervoort, M., M. Crozatier, D. Valle, and A. Vincent. 1999. 'The COE transcription factor Collier is a mediator of short-range Hedgehog-induced patterning of the *Drosophila* wing', *Curr Biol*, 9: 632-9.

Wenick, A. S., and O. Hobert. 2004. 'Genomic cis-regulatory architecture and trans-acting regulators of a single interneuron-specific gene battery in *C. elegans*,' *Dev Cell*, 6: 757-70.

Appendix

7.1 Summary

In this thesis, we studied two key challenges that cells face throughout their lifetime: the correct timing of developmental events during development, and second, when cells have finally reached their terminal cell identity, this cell identity has to be maintained. During development, cells proliferate, cell fates are specified and cells differentiate into their terminal cell fate, cells move towards their end location, and they grow in volume. It is crucial that all those developmental events are executed at the right time and in the correct sequence, in order to prevent malformation and dysfunction of tissue which could lead to incompatibility with survival. This correct timing of developmental events is the first key challenge that we will study in this thesis, in which we focus on the cell growth. The second challenge that we will study is one that cells face when they have reached their terminal cell identity, after which the terminal cell identity has to be maintained during the life-span of the cell.

The quantification of expression levels of genes of interest has been the main driver for studying the research questions in this thesis. To quantify expression levels of genes of interest, we mainly used single molecule FISH (smFISH) in fixed animals. The highly precise measurements of gene expression enabled by smFISH were crucial for quantification of terminal selector genes (Chapter 3) and oscillatory gene expression (Chapters 4 and 5). In **Chapter 2**, we describe the analysis software package that we wrote to analyze smFISH data sets. It is specifically aimed at smFISH data obtained with *C. elegans* experiments, which comes with specific challenges that have not yet been addressed with existing smFISH analysis packages for smFISH data. In contrast to most existing smFISH analysis packages that focus on single cell data, *C. elegans* data requires a workflow suited for analyzing large sets of smFISH microscopy images in which only a few cells are annotated within the animals. We optimized the workflow, to enhance the efficiency and to reduce the time spend on analysis of large *C. elegans* data sets. In addition, we added an annotation tool to the package that allows for easy annotation of the developmental stage of each individual animal. This makes it possible to directly link smFISH measurements with the developmental stage of the individual animals, aiding in studying temporal developmental related dynamics in expression.

Many cells in the human body are renewed at timescales from days to several years. However, neuronal cells are hardly renewed and those neuronal cells face the

challenge of maintaining their specialized functions over the entire lifespan. How the terminal cell identity is maintained over such long timescales despite intrinsic molecular fluctuations, is not known. In **Chapter 3**, we explore the underlying mechanism behind the maintenance of terminal cell identity by a genetic cell fate switch. Here, we focus on the zinc-finger transcription factor *che-1* in the ASE neurons, required for sensing small soluble molecules in the animals' environment, including chemotaxis to salts. CHE-1 is required for the induction and maintenance of the ASE cell fate, by inducing the expression of hundreds of target genes required for the specialized function of the ASE neurons. Crucially, it also induces its own expression via positive autoregulation, acting as a genetic cell fate switch. Genetic switches are known to be vulnerable to intrinsic molecular fluctuations, causing spontaneous switching between stable states, which in case of the *che-1* switch could lead to spontaneous loss of expression of the CHE-1 target genes. We identified a novel mechanism which ensures the life-long maintenance of ASE neuron fate by the *che-1* switch. We measured the copy numbers and lifetimes of *che-1* mRNA and CHE-1 protein, and combined these measured parameters with stochastic simulations. Those simulations predicted that continued *che-1* expression is ensured by preferential binding of CHE-1 to the *che-1* promoter. We validated this mechanism by transient depletion of CHE-1 levels in the ASE neurons, showing resilient expression of *che-1* while the expression of the targets vanished. Additionally, we identified a 130 bp fragment in the *che-1* promoter region responsible for the stable *che-1* expression. Within this fragment, we identified an *Otx*-related homeodomain binding site which confers stability of *che-1* expression. The network architecture of a genetic switch like *che-1* is highly conserved among organisms. Subsequently, the mechanism we uncovered might explain how terminal identity is maintained over the lifetime of other cells and organisms.

In **Chapter 4**, we set out to validate two clusters that we found with a novel percolation clustering approach on single-animal RNA-sequencing data on young adult animals developed Werner *et al.* (2020). Moreover, genes within the clusters, but also between the clusters, showed correlated variation in the RNA-sequencing data. Interestingly, one cluster consists out of genes specifically expressed in the pharynx, while the other cluster consist out of genes specifically expressed in neurons. We also attempted to identify a possible origin underlying the correlated variation in expression between genes in these two seemingly independent clusters. The smFISH method enabled us to quantify the mRNA levels of genes in the clusters more precisely than the RNA-sequencing method, and the method also provides spatial expression information. We quantified the expression levels of gene pairs of the two clusters in animals with a similar age as in the RNA-sequencing data, showing that indeed at least one of the clusters consists of genes

that are correlated in expression, validating one cluster. The results showed that the correlated expression of the genes in the validated cluster likely originates from the oscillations in gene expression related to development. Previous findings proposed a possible origin underlying the correlated variation in which transgenerational inheritance of starvation history plays a role. However, our results on animals that have starved ancestors showed no increased correlation between or within the clusters. Interestingly, we did find decreased expression levels of TGF- β ligand *daf-7* in animals with starved ancestors, a gene which has been shown in previous studies to be related to the general starvation response of animals and to transgenerational inheritance. Another origin of the correlated variation could be a starvation response, since animals in the RNA-sequencing experiment were starved between 45-60 minutes before they were fixed for sequencing. We found that this induction of starvation enhanced the correlation between the selected genes of both clusters, by inducing in one cluster a developmental stage-specific expression pattern that resembled that of the other cluster.

Before cells are terminally differentiated and functioning within the organisms, they go through many developmental events, including specification, differentiation, proliferation, movement and growth. All those events are series of sequential events that need to be executed at the correct time and sequence. This process requires time-keeping of the developmental events, which is a key challenge that we addressed in **Chapter 5**. Recent studies have shown that *C. elegans* has a potential developmental clock driven by oscillations in gene expression, that is involved in the time-keeping of the periodic developmental process called the molting cycle, in which the skin of the animals is renewed and shed each stage of the development. So far, it remained elusive if it could also be involved in other developmental events. In Chapter 5, we focused on the pharynx of *C. elegans*, the feeding organ of the animal. We showed that a group of genes that have oscillating expression with a peak phase outside the molting period, consists out of genes related to growth of the pharynx. By following the growth of the pharynx in time, we showed that the growth of the pharynx has a cyclical character. Using smFISH to quantify expression levels, we found that the oscillating expression of the selected genes appeared to be coordinated with the growth dynamics of the pharynx in development. By arresting the development of the animals through starvation, we showed that when the pharynx is halted in its growth, the oscillating gene expression is also halted at a specific phase corresponding to the developmental checkpoints. These observations suggest that the oscillating gene expression in the pharynx is functionally coupled to the growth of the organ.

7.2 Samenvatting

In dit proefschrift zijn twee belangrijke uitdagingen behandeld die een cel tegenkomt tijdens zijn levensduur: het correct timen van de verschillende ontwikkelingsprocessen die de cel ondergaat tijdens zijn ontwikkeling, en het behouden van zijn terminale identiteit nadat deze bereikt is. Tijdens de ontwikkeling zijn cellen onderhevig aan proliferatie, specificatie en differentiatie, verplaatsing en ook groei van de cellen in volume. Het is cruciaal dat deze ontwikkelingsprocessen op de juiste tijd en in de correcte volgorde worden uitgevoerd, om te voorkomen dat weefsel misvormd raakt of zelfs niet goed meer kan functioneren wat kan leiden tot het overlijden van het organisme. Het correct timen van de verschillende ontwikkelingsprocessen is een uitdaging die wij in dit proefschrift bestuderen, waarbij wij focussen op de cel groei. De andere uitdaging die cellen tegenkomen wanneer zij hun ontwikkeling zo goed als hebben afgerond, is het behoud van de terminale cel identiteit tijdens de rest van hun levensduur.

Het kwantificeren van de expressie levels van relevante genen is een belangrijke drijfveer geweest in dit proefschrift voor het beantwoorden van de gestelde onderzoeksvragen. Om de expressie levels te kwantificeren, hebben wij gebruik gemaakt van de single molecule FISH (smFISH) methode, welke plaats vindt op gefixeerde wormen. Doormiddel van de smFISH methode waren wij in staat om met hoge precisie expressie levels te meten, die cruciaal waren voor de kwantificatie van expressie levels van terminale selector genen (Hoofdstuk 3), en voor de kwantificatie van oscillerende expressie van genen (Hoofdstuk 4 en 5). In **Hoofdstuk 2** beschrijven wij ons softwarepakket welke is ontwikkeld voor het analyseren van smFISH data, en welke is geschreven in Python. Het softwarepakket is specifiek gemaakt voor het analyseren van smFISH data gegenereerd door *C. elegans* experimenten. Deze data komt namelijk met specifieke uitdagingen die nog niet aangepakt zijn door bestaande softwarepakketten. In tegenstelling tot de meeste bestaande softwarepakketten voor smFISH analyse, die gefocust zijn op het analyseren van single-cell data, vereist de analyse van smFISH datasets van *C. elegans* een andere aanpak. Hierbij moet de aanpak meer gericht zijn op het analyseren van grotere datasets met daarin microscoop plaatjes van individuele wormen waarin hooguit enkele cellen zitten die geannoteerd moeten worden. Wij hebben de workflow van de analyse geoptimaliseerd voor de grote datasets van *C. elegans*, zodat het proces efficiënter verloopt en de benodigde tijd voor de analyse wordt gereduceerd. Daarnaast hebben wij ook een annotatie tool toegevoegd aan het softwarepakket, waarbij de gebruiker op een eenvoudige manier elke individuele worm kan annoteren met het daarbij horende ontwikkelingsstadium. Deze tool maakt het mogelijk de smFISH metingen direct te linken met de geannoteerde

ontwikkelingsstadia, waarmee ontwikkelingsdynamica in de genexpressie van *C. elegans* bestudeert kan worden.

Veel cellen in het menselijk lichaam worden regelmatig vervangen, met een tijdschaal van enkele dagen tot meerdere jaren. Daarentegen worden neuronale cellen bijna niet vervangen, waardoor deze cellen de uitdaging hebben om een levellang hun gespecialiseerde cel functies te behouden. Hoe de terminal cel identiteit van deze cellen wordt behouden over zulke lange tijdschalen is niet bekend. In **Hoofdstuk 3** onderzoeken wij het onderliggende mechanisme dat ten grondslag ligt voor het behoudt van de terminale identiteit door een genetische switch. Wij focussen hier op de zinc-finger transcriptie factor *che-1*, die tot expressie komt in de ASE neuronen. Deze neuronen zijn verantwoordelijk voor de detectie van kleine oplosbare moleculen in de omgeving van een worm, waaronder zouten. CHE-1 is noodzakelijk voor de inductie en het onderhouden van de identiteit van de ASE neuronen. Hierbij induceert CHE-1 de expressie van honderden target genen die verantwoordelijk zijn voor de gespecialiseerde functie van de ASE neuronen. Essentieel hierin is dat CHE-1 ook zijn eigen expressie induceert door middel van positieve autoregulatie, waardoor het de architectuur krijgt van een genetische switch. Het is bekend dat genetische switches gevoelig zijn voor intrinsieke moleculaire fluctuaties, wat kan leiden tot een spontane switch tussen stabiele staten. In het geval van de *che-1* switch zou dit kunnen leiden tot het spontaan verliezen van expressie van de *che-1* target genen. In Hoofdstuk 3 hebben wij een nieuw mechanisme geïdentificeerd dat het onderhouden van de cel identiteit van de ASE neuronen over lange tijdspannen kan waarborgen door middel van de *che-1* genetische switch. Wij hebben de aantallen en de levensduur van *che-1* mRNA en eiwit gemeten, en wij hebben deze gecombineerd met onze stochastische simulaties van de *che-1* switch. Deze simulaties voorspelde dat *che-1* expressie wordt behouden door middel van een preferentie in de binding van CHE-1 op zijn eigen promotor. Wij hebben dit kunnen valideren door CHE-1 levels tijdelijk te verlagen in de ASE neuronen, waardoor wij konden laten zien dat de expressie van *che-1* haast niet veranderd terwijl de expressie van de target genen bijna verdween. Een 130 bp fragment in de *che-1* promotor bleek verantwoordelijk te zijn voor deze stabiele expressie van *che-1*. Daarnaast bevindt zich in dit fragment een bindplaats voor een *Otx*-related homeodomein eiwit, welke helpt bij de stabiliteit van de expressie van *che-1*. De architectuur van het netwerk van de *che-1* switch is geconserveerd binnen de verschillende organismen. Dit suggereert dat het mechanisme wat wij gevonden hebben mogelijk kan verklaren hoe terminale cel identiteit behouden blijft over lange tijdsduren in andere cellen of organismen.

In **Hoofdstuk 4** was het de bedoeling om twee cluster te valideren die gevonden waren met een nieuwe clusteringsaanpak ontwikkeld door Werner *et al.*

(2020), waarin gebruikt wordt gemaakt van de percolatie theorie, toegepast op RNA-sequencing data van individuele jongvolwassen wormen. De genen in de clusters en de genen tussen de clusters, lieten in beide gevallen correlatie in genexpressie variatie zien. Deze correlatie is interessant gezien het verschil in expressie patroon tussen de clusters, waarbij cluster 1 specifiek tot expressie komt in de pharynx en cluster 2 in neuronen. Wij hebben ook geprobeerd de mogelijke biologische oorsprong onderliggend aan de correlatie tussen de clusters te achterhalen. Door middel van de smFISH methode hebben wij de mRNA levels van genen in de clusters gekwantificeerd, en daarnaast ook de gedetailleerde spatiale expressie informatie bestudeerd. De expressie levels van geselecteerde genen paren van de twee clusters werden gekwantificeerd in wormen met dezelfde leeftijd als in de RNA-sequencing data. Hieruit bleek dat de genen binnen in één cluster gecorreleerde expressie vertoonde, waardoor dat cluster gevalideerd kon worden. De smFISH data van dit gevalideerde cluster liet zien dat de correlaties waarschijnlijk voortkomen uit oscillaties in genexpressie die gerelateerd zijn aan de ontwikkeling. Bevindingen uit eerder onderzoek laten zien dat de mogelijke onderliggende oorsprong van de genexpressie variatie in de clusters, ook voort kan komen uit transgenerationale overerving van in het verleden ervaren uithongering. Echter laten onze smFISH resultaten op wormen met voorouders die uitgehongerd zijn geen toename zien in correlatie tussen genen binnen de clusters of tussen de clusters. Wat wij wel konden laten zien is dat de TGF- β ligand *daf-7* lagere expressie levels heeft in wormen met voorouders die uitgehongerd zijn ten opzichte van de controlegroep. Het gen *daf-7* is in andere studies al eerder in verband gebracht met uithongering en transgenerationale overerving. Een andere mogelijke onderliggende reden voor de genexpressie variatie in de clusters is uithongering, omdat de wormen tijdens de RNA-sequencing experimenten uitgehongerd werden voor 45-60 minuten voordat de wormen werden gefixeerd. Wij lieten zien dat de reactie op uithongering in de wormen de correlatie in genexpressie tussen de twee clusters versterkte. Dit komt doordat één van de clusters verandering in genexpressie laat zien die afhankelijk is van het ontwikkelingsstadium van de worm, waardoor het temporale patroon in genexpressie overeenkomt met die van het andere cluster.

Cellen ondergaan vele ontwikkelingsprocessen voordat ze terminaal gedifferentieerd zijn en functioneren in weefsel van organismes, waaronder specificatie, differentiatie, proliferatie, lokalisatie en groei. Al deze ontwikkelingsprocessen zijn series van sequentiële gebeurtenissen die plaats moeten vinden in de juiste volgorde en op het correcte moment. De juiste timing van de gebeurtenissen is hierin essentieel, welke een belangrijke uitdaging is tijdens de ontwikkeling van cellen die wij behandelen in **Hoofdstuk 5**. Uit recente studies blijkt dat *C. elegans* een potentiële ontwikkelingsklok bezit die gedreven is door oscillaties in genexpressie en betrokken is bij de correcte uitvoering van ontwikkelingsprocessen

in de tijd. Deze potentiële ontwikkelingsklok lijkt voornamelijk betrokken te zijn bij de periodieke vervelling van de huid van de worm tijdens de verschillende postnatale ontwikkelingsstadia. Het is niet bekend of de oscillaties in genexpressie ook betrokken zijn bij andere ontwikkelingsprocessen. In Hoofdstuk 5 focussen wij op de pharynx van *C. elegans*, een orgaan is dat verantwoordelijk is voor het naar binnen pompen van eten. Wij hebben laten zien dat een groep met genen die oscilleert met een fase die net na de periode valt waarin de vervelling plaats vindt, bestaat uit genen die gerelateerd zijn aan de groei van de pharynx. Door het volgen van de groei van de pharynx in de tijd, hebben wij kunnen aantonen dat deze groei een cyclisch karakter heeft. Met behulp van smFISH konden wij de oscillaties in genexpressie aantonen van de geselecteerde genen, en daarnaast konden wij laten zien dat de oscillerende genexpressie en de groei dynamica van de pharynx gecoördineerd zijn tijdens de ontwikkeling. Door de wormen uit te hongeren stopt de ontwikkeling, waarbij ook de pharynx stopt met groeien, daarbij zagen wij dat de oscillerende genexpressie van de geselecteerde genen ook stopten in een specifieke fase die overeenkwam met ontwikkelingscheckpoints. Deze observaties suggereren dat de oscillerende genexpressie in de pharynx functioneel gekoppeld is met de groei van het orgaan.

7.3 Acknowledgements

Since you are reading this, which is most likely one of the first chapters you read after receiving my dissertations (and perhaps even one of the few), you are probably eager to read whom I am thanking. I have read acknowledgements of dissertations saying that this part is something they look forward to writing. For me it is the opposite, I am scared that I forget to acknowledge someone and to let someone down. So here is a warning up front, if your name is not in my acknowledgements, you are most likely somewhere in a cherished place in the back of my mind while you were a part of my PhD journey, even if it was a small one.

First off, I want to thank my supervisor Jeroen for giving me the opportunity of doing a PhD in his group. The first years of my PhD weren't the easiest, but it definitely shaped me, in a good way. You gave me the freedom to find my own path during my PhD journey. I want to thank you for your trust, and for everything you taught me scientifically and also personally. I truly admire your passion for science, and your persistence and optimism when something does not turn out the way we hoped it would. And I will surely miss those funny random conversations during lunch with your inexhaustible ability to come up with cool and quirky science facts.

Erwin, Stefan, Rik, Pieter Rein and Lidewij, thank you for being part of my committee, and for taking the time to read and evaluate my dissertation. I look forward to our scientific discussions.

Next, I want to thank all the current members of the Quantitative Developmental Biology group. Yvonne, thanks for always helping us out in the lab, with the general lab stuff of course, but also, with our more random Friday afternoon experiments, such as when we glued single worms to wells, which sort of worked in the end ;) And thank you for being a listening ear when I needed one. Rutger and Xuan, thanks for the useful discussions and nice conversations. I always admired your work in those delicate organoids, where you traced fixed organoids back in time or shot high power lasers at them, keep up the good work! Timo, we only overlapped for (not even) one year, but it feels like way longer. I admire your tremendous enthusiasm, and your tireless love for cooking, books and plants. I am thankful for the scientific but also non-scientific discussions that we had, and for the contributions that you made to the pharynx project. Burak, I really enjoyed having you as my office mate. We definitely shared the same sense of humor, which could be a bit dark from time to time. Thank you for bringing joy into our little office with witty and also sometimes more philosophical conversations.

Furthermore, I want to thank the former lab members and the several of the

master/bachelor students that visited the lab. Jason, I am thankful for your support during my PhD, and the nice conversations we had about life and science, sometimes accompanied by an exotic beer. If you ever happen to setup your own beer/cider label, I would definitely be a loyal costumer of yours. Olga, we went together on an amazing trip to a summer school and conference in the US. I enjoyed your company, and how optimistic and genuine you are. I wish you all the best in Austria, together with Mario and Ariadna. Guizela, I always liked our funny conversations about our shared love for cats during lunch breaks. Also, thank you for your (digital) support the last few months, it helped me gain more confidence. I also want to thank Nicola and Simone, for their valuable input during my PhD. Last, I would like to thank Dimphy for her work during her internship, which was not the easiest or straightforward project. And Simon, who is also my former flat mate, I really enjoyed your company during your short internship. I hope that one day your preliminary organoid model can be used for creating cool simulations.

Many thanks to my collaborators in Rotterdam. Ser, I am very happy with our collaboration on the *che-1* project, it would not have been the same without your help. I very much enjoyed our scientific discussions and how we were able to come up with new cool ideas and experiments. I wish you all the best in Cambridge, together with Tintin and Luc. Gert, as with Ser, I am very grateful for our collaboration. Thank you for your pragmatic and down to earth approach (*zonder toeters of bellen*), that was very much appreciated during our discussions. Suzanne, thank you for your help, with the immunostaining and also recently for making a new line, and for the nice conversations that we had when I was at Erasmus.

Next, I want to acknowledge all the members of the FOM consortium, for the interesting scientific discussions during our annual meetings. In particular, I want to thank Mathijs and Steffen, we all worked on the same RNAseq project, and we also endeavored into some cool behavioral experiments, which in the end appeared to be way more complicated than we anticipated. Nonetheless, I very much enjoyed our discussions and sharing the feeling of excitement and curiosity when doing new experiments. Mathijs, I envy the amount of inexhaustible energy and positivity you have. Thank you for your support, especially at the end (*de laatste loodjes*). Steffen, all the best with your new position at the WUR! I am curious to see what kind of research directions you will take, I am certain it will be cool. Tom, I am thankful for the many scientific discussions we have had over the course of my PhD.

In the first years of my PhD, I shared the lovely, albeit sometimes chaotic office with Simone, Agata and Michele, whom were always able to bring a smile on my face. I also want to acknowledge other PhD students/Postdocs/technicians from AMOLF, including Simone B, Harmen, Celine, Federica, Mareike, and many other that I crossed paths with. And not to forget, many thanks to the support team at

AMOLF, including Marko, Jan Bonne, Hincó, Clyde.

Daarnaast wil ik iedereen bedanken van buiten de academia. Allereerst wil ik Aagje en Iris bedanken, twee lieve vriendinnen die ik ken van mijn studententijd, van toen we nog op de IBB woonde in onze oude studentenflat (NUMMER 169! ;). Onze gezamenlijke etentjes zijn nu echt een gemis geworden sinds covid. Het was heerlijk om even te kunnen lachen en relaxen met z'n drieën, wat zeker tijdens de wat zwaardere tijden van mijn PhD zeer welkom was. Hopelijk zien we elkaar in echt snel weer. Eva, Wesley en Amir, bedankt voor de fijne etentjes en game avonden. Eva, ik ben je erg dankbaar voor je steun tijdens mijn PhD, zeker als het ging om de herkenbare PhD struggles die we samen deelde. Ik kon je altijd even bellen of appen als ik ergens mee zat. En de leuke uitjes die we samen ondernamen om even te kunnen ontspannen. Heel veel succes met je laatste loodjes, je kan het! En natuurlijk Lichelle, wij kennen elkaar nu al bijna 26 jaar volgens mij? Ik ben ontzettend dankbaar voor onze vriendschap, en voor de steun die ik aan je heb gehad. Ik vind het erg knap wat je allemaal voor elkaar hebt gebokst, zeker nu je in je laatste jaar van je PhD zit en die kleine knappe Noah op de wereld hebt gezet. Ik hoop je snel weer te zien, samen met Noah en Yorrick erbij.

En als laatste wil ik mijn familie bedanken. Mam, bedankt dat je altijd voor mij klaar stond. Ik kon je altijd bellen als er iets was, of het nu iets groots of kleins was. Of om gewoon “even” (al liep dit altijd uit op meer dan een uur) bij te kletsen. Hopelijk kunnen we snel weer een keer samen ouderwets naar de Intratuin. Pap, jij ook bedankt dat ik altijd bij je terecht kon voor advies, of ik nu een luisterend oor nodig had of een vraag over mijn autootje, je stond altijd voor mij klaar. Ik ben jou en mam ook heel erg dankbaar voor de kansen die jullie mij en Thom gegeven hebben. Thom, waar moet ik beginnen, ik ben erg dankbaar jij mijn kleine broertje bent (al ben je niet zo klein meer). Dankje dat ik altijd bij je terecht kon. Ik heb genoten van onze avonturen naar al die verschillende musea, lange wandeltochten en fotografie uitstapjes, en ik hoop dat er nog vele bijkomen.

